Appendix B

A Model of Lifetime Earnings Patterns [1]

1. Introduction

Social security law bases benefits on an average of the best years of earnings of an individual worker. Eventually retirement benefits will be based on the 35 years of highest earnings. This Panel has endorsed the principle of lifetime average earnings and recommends the eventual use of a 35-year average of indexed earnings. At present the averaging period is considerably shorter and no earnings before 1951 enter the calculation of benefits for most workers. To understand the future shape of the social security program and to have a model for cost estimation, it is thus necessary to have some understanding of the patterns of earnings over workers' entire lifetimes. No body of data exists which reports on the earnings of a large number of workers over full working lifetimes. Hence we have undertaken to estimate a model of lifetime earnings based on a large body of earnings data reported to the Social Security Administration since 1956.

At the start of this project, the 0.1 percent Continuous Work History Sample containing estimated [2] earnings for 1956 to 1971 was available. In addition the data for 1972 were available except for the level of self-employment earnings. [3] Since the primary purpose of the model was to project earnings histories into the future, we have fitted the model only to male earnings, given the belief that future female earnings are likely to differ sharply from those of the past. [4] The task was to move from this set of data containing up to 16 observations per person to a model giving the distribution, not just the average of lifetime earnings patterns. [5]

The model described below was used for simulations of wage histories which were used to project retirement benefits, yielding estimates in a form which could readily be incorporated into the long-run cost estimation procedure of the Office of the Actuary. An important conclusion of the simulation study is that cost estimates depend significantly on the specification of the random component of earnings growth as well as depending on the typical age structure of individual earnings.

In addition to being a basis for simulations, the model developed yielded a number of conclusions on the patterns of male earnings experienced over the time period analyzed, confirming the statistical findings described in Chapter 6 and Appendix A. Typically, until age 35 individuals experience wage growth that is much more rapid than the growth of average earnings in the economy. Between ages 35 and 64 individual earnings growth does not differ too much

[1] This Appendix is based on the joint research of Peter Diamond, Richard Anderson, and Yves Balcer. The basic model was developed by Roger Gordon in his Ph.D. dissertation at MIT and adapted by him for Social Security data. Jerry Hausman has contributed a great deal of econometric advice. The calculations could not have been performed without the assistance of the Social Security Administration, especially Aaron Prero, Barry Bye, and John Spencer. Helpful suggestions have been made by a large number of others. Responsibility for errors and the like remain with Diamond, Anderson, and Balcer.

2 We have used the Method II estimate which extrapolates earnings (separately by employer) for the remaining quarters of the year for any employee whose reported earnings reach the taxable maximum. In addition no estimate is available for self-employment income of those who earn above the maximum as employees.

3 But we did have an indicator of whether self-employment earnings existed.

4 We chose to make no use of data on location and industry of employer available starting in 1971.

5 Earnings outside covered employment (e.g., for the U.S. government) are not reported. Thus we have zeros in the data both for people without earnings and for those working in uncovered jobs.

from the growth of the economy-wide averages for those who do not claim retirment benefits. There are large unexplained elements in individual earnings after one has adjusted for the typical age structure and for other components of steady growth. Adjusted for movements out of covered employment, the typical age structure of earnings does not vary much with the level of earnings between the upper two-thirds of the income distribution. It is different at the bottom of the income distribution showing a less rapid growth to the level of peak earnings. The random component in earnings is smaller in percentage terms the higher the income level.

2. Framework of Analysis

Ideally one would want to explore the determinants of earnings levels for different workers. This would imply an examination of the demand for and supply of labor of different ages, skills, experience levels, etc. Such an approach seemed considerably beyond the capabilities of this study. Thus we have taken the lesser task of examining the data on wages in the period 1956-71 in order to select a pattern of lifetime wages which is consistent with the observed pattern and a suitable extension to cover entire lifetimes. Restating this perspective, an individual's history can be considered as a random draw from some distribution defined over a 45 dimensional random vector representing annual earnings from ages 20 to 64. Given the outcome of this random draw, the highest 35 earnings in the single draw are selected to determine the average earnings of a particular worker. The problem is to describe the distribution.

If the distribution were believed to be multivariate normal, one could directly consider the 45 dimensional vector and estimate means, variances, and covariances where age differences were not too large.[6] A complete distribution could then be constructed by extrapolating the variance-covariance matrix to the unobserved off-diagonal terms. However, the distribution is very far from being multivariate normal .[7] Not knowing any suitable way to move from a variance-covariance matrix plus marginal distributions to either a full description of the distribution or to the needed order statistic (the mean of the 35 largest earnings), we have followed the route of making assumptions on lifetime patterns which lead to ordinary least squares regressions and an estimation of the distribution based on regression coefficients and the distribution of residuals.

3. Model 8

Before considering the structure of the model, let us detail the earnings measure to be described. To avoid the issue of explaining both inflation and productivity growth, it seems appropriate to relate earnings of individuals in a particular year to average earnings in that year. There are several different average earnings series which might be used for this indexing purpose. It is not clear that there is a particularly correct index to use, in the absence of a theory of the impact of inflation, productivity gains and the age and sex mix of the labor force on the age structure of earnings. If one assumed no effects from these

6 An estimate of the variance-covariance matrix is being calculated as an evaluation of the estimates developed below. The calculation was not ready in time to be included here.

7 To examine normality in the distribution of earnings growth, five birth cohorts (1907-1911) were examined for two pairs of years. For each pair of years the logarithm of the ratio of estimated earnings in t+ I to estimated earnings in t was calculated for each worker with positive earnings in all three of t-1, t, and t+1. Then the distributions were calculated. In addition each cohort was divided into thirds by income in t-1 and the procedure repeated for each third. The distributions were consistently different from the normal distribution. The coefficients of skewness were mostly negative and generally less than -1. The coefficients of kurtosis were all positive and almost all larger than 10 and one-third larger than 20. The standard deviations were generally between 1/3 and 2/3.

8 For a fuller description of this model and another use, see Chapter III of the unpublished MIT Ph.D. dissertation of Roger Gordon, "Essays on the Causes and Equitable Treatment of Differences in Earnings and Abilities." The model there was adapted for this problem by Gordon.

variables, a fixed weight average of earnings of different ages would be the correct measure. However, the analysis here uses a wage index constructed from includable[9] observations for all males age 20 and over who were in the smaller sample.[10] The average wage series is shown in table 1, along with the economy-wide average estimated covered earnings and average covered first quarter wages and salaries, the latter being the index used to increase the maximum taxable earnings base.

TABLE 1.—MEAN WAGE INDEXES

Date	Annual earnings males 20 and over in sample	Estimated annual covered earnings	1st quarter wages and salaries	Annual earnings males 25 to 64 in sample
1956	$4,076	$3,207	$879	$4,638
1957	4,100	3,314	927	4,764
1958	4,177	3,390	957	4,805
1959	4,500	3,557	989	5,166
1960	4,693	3,656	1,032	5,428
1961	4,766	3,720	1,064	5,514
1962	4,899	3,890	1,109	5,734
1963	5,016	4,002	1,136	5,907
1964	5,223	4,191	1,171	6,232
1965	5,542	4,359	1,189	6,696
1966	5,963	4,618	1,241	7,335
1967	6,151	4,852	1,320	7,595
1968	6,565	5,147	1,413	8,175
1969	7,045	5,453	1,486	8,787
1970	7,530	5,733	1,563	9,396
1971	7,863	6,013	1,658	9,833
1972	8,098	6,399	1,802	10,244

To test the importance of the choice of index, the basic equation was re-estimated using another index shown as column five in the table. When the coefficients are adjusted for the more rapid growth of the alternative average earnings series (approximately 0.5 percent per year average) they are essentially the same.

Given the complexities (and lack of importance for these purposes) associated with earnings of the young, no earnings before age 20 are considered. In addition no attempt was made to estimate earnings of those over 64. The presence of social security makes the determinants of the earnings of the elderly (primarily the retirement test) somewhat different from those of younger workers. With the need to register for medicare benefits, registration for social security benefits is not a useful indicator of partial retirement for those over 65 for much of the data period. The expectation that the random structure of the model is more likely to be multiplicative than additive led to a formulation in logarithms.[11] Thus the variable to be explained is defined as

$$W_t^h = \log\left(\frac{\text{earnings of person } h \text{ in year } t}{\text{average earnings in year } t}\right) \tag{1}$$

In addition W^h is defined as the average of the W_t^h taken over the years when earnings are positive.

A problem inevitably arises in treating years when earnings are zero. It was decided not to attempt to simultaneously estimate the probability of zero and the distribution of earnings when positive, but to proceed on the assumption that the two parts are separable,[12] treating all zeros as missing observations.[13]

[9] The definition of the set of observations included in the analysis will be given below.

[10] The index was calculated using approximately twice as many persons as were used in the regressions.

[11] No attempt was made to examine whether some other transformation of earnings was a more appropriate one to use in a linear regression.

[12] For a rudimentary model of the probability of a zero, see section 12, below. For the relative frequency of zeros by income level, see section 7.

[13] In addition, with death or receipt of retirement benefits during a year, the earnings of that year or any later year were eliminated from the sample. Earnings in years with receipt of disability benefits were also eliminated. With retirement late in a year, this would be the procedure to evaluate benefits on retirement but not necessarily on recomputation a year later.

That is, earnings are estimated conditional on being positive. The assumption for estimation purposes is that for an individual the probability of a zero is independent of the earnings record which would occur in the absence of zeros, although it may vary with age and permanent characteristics of an individual. No further adjustments are made for these missing observations, since the procedure followed is unbiased and while such adjustments would affect efficiency, the sample is quite large. Before examining further refinements made to adjust for the presence of zeros in neighboring or previous years, let us consider the basic model relating this earnings variable to age.

The basic assumption of the model is that the path of expected values of W^h_t has the same slope for all people, but with different heights for different people. That is, in a log wage-time diagram all people follow parallel paths, randomness aside, but intercepts differ across individuals. The assumption that the steepness of income growth paths does not vary significantly with income level may seem surprising to some. Some support for this assumption except at low incomes was described in Appendix A. Below in sections 7 and 9 we will consider further evidence that this assumption is a reasonable one, for all but the lowest income level. We will also consider a modification of the model to allow for systematically different individual paths, although the modification was not pursued very far. To express the model formally let us define a set of age variables A^h_{it}

$$A^h_{it} \equiv \begin{cases} 1 & \text{if person } h \text{ becomes } i \text{ years old in year } t \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Then the basic model is

$$W^h_t = a^h + \sum_{i=20}^{64} b_i A^h_{it} + u^h_t \tag{3}$$

where a^h is the coefficient on an individual dummy[14] and u^h_t is a random variable with zero mean and finite variance. The problem is to estimate the distribution of a^h, the coefficients b_i, and the distribution of random errors u^h_t.

The procedure is to pool all the W^h_t for all people and all years in a single ordinary least squares regression. There are two basic assumptions underlying this formulation: first, that the expected path of log earnings has the same slope for all people, second, that the individual characteristics which determine the height of the path stay constant over a lifetime. The slope assumption will be discussed further below. To assume a lifelong individual constant is to assume that all deviations from the trend are captured in the structure of the random elements u^h_t in the wage equation (3). The two structures examined are u^h_t independent random variables[15] and u^h_t having a first order autocorrelation structure

$$u^h_t = \rho u^h_{t-1} + v^h_t \tag{4}$$

where v^h_t are independent random variables. Given the absence of explanatory variables other than age and presence in covered employment, this random structure does not seem adequate to capture large changes in general earnings, whether through changes in earning ability (e.g., health) or taste. In particular it might be interesting to explore a model where the individual constants could

[14] An individual dummy is 1 for the wage observations of the particular individual and zero otherwise.

[15] Not necessarily identically distributed for different ages.

change withing a lifetime.[16] Since the model is fitted to a 16-year period and then used for simulation over a 45-year period, this misspecification probably involves too few large changes within a lifetime and too much short period noise as the random elements attempt to capture both of these effects.

4. Age Structure Variables

To directly employ equation (3) on a large body of data would not be appropriate since there would be an inconvenient number of right hand side variables—45 plus the number of people in the sample. The procedure actually followed was that of subtracting the means of all variables for each person from the values of the variables. Thus the equation fitted became

$$W_t^h - \bar{W}^h = \sum_{i=20}^{64} b_i (A_{it}^h - \bar{A}_i^h) + u_t^h \tag{5}$$

Since this equation would give too many coefficients to be easily handled, for the ages 20 to 59, they are constrained to be piecewise linear in 5-year intervals.[17] The procedure is to define 9 dummy variables A'^h_{jt} defined over the values (0, 0.2, 0.4, 0.6, 0.8, 1) reflecting the five-year intervals between 20 and 60. An individual whose age is a multiple of 5 in a year would have the appropriate dummy set equal to one, all other dummies being zero. For a year when a person's age is not a multiple of five, two dummies, representing the neighboring multiples of five are nonzero, with the weights (adding to one) such that his age is a weighted average of the two five-year points. Thus a 22-year-old has A'_1 equal to 0.6, A'_2 equal to 0.4 and all other dummies set equal to zero. Thus the fitted equation became

$$W_t^h - \bar{W}^h = \sum_{j=1}^{9} b'_j (A'^h_{jt} - A'^h_j) + \sum_{i=61}^{64} b_i (A_{it}^h - A_i^h) + u_t^h \tag{6}$$

Because the complete set of age variables display perfect collinearity, the dummy for age 50 is omitted in the regression. Hence coefficients measure the difference between the coefficient for some other age and that for age 50.

The equation was fitted to two bodies of data—a subsample of the 0.1 percent CWHS of 1,576 persons (16,747 observations), on which we tried out different models and tested some ideas, and the entire 0.1 percent CWHS of 65,119 persons (689,377 observations). The results for this equation are reported in tables 3 and 4 and discussed in section 6 below. Given the large size of the samples, in the estimation no adjustments are made for heteroskedasticity or autocorrelation of u_t^h.

5. Dummy Variables for Noncovered Employment

The formulation in the previous section makes no use of the available information on the absence of all covered earnings in some years. In addition, consideration of the presence of a zero in the earnings history together with the method of estimating earnings for this data set indicate an error in the data that requires further adjustment. Let us start with the use to be made of zeros in an earnings record.

As indicated at the start, a separate model is being estimated to yield the probability of positive earnings in a year. In wage simulation, one then combines a simulation of positive earnings with a probability of zeros in the earnings

[16] The importance of changes in the individual constant could be tested somewhat by examining earnings predictions for 1972 using different length periods to estimate the individual constants (but the same age structure of earnings). If the individual constants are stable, the longer the time period used in their estimation the better the estimate. If they are not stable, use only of recent years might give a better estimate.

[17] In retrospect, the ages 20 to 25 should also have been fitted separately since the growth rate seems to vary considerably between those years.

record. Thus if the presence of a zero does affect earnings levels in other years, it would be appropriate to include such an effect in the simulation. Most commonly, one would expect a zero in an earnings record to represent employment in noncovered employment.[18] In addition, some zeros result from unemployment of long duration or withdrawal from the labor force.[19] It is unlikely that such departures from covered employment exactly coincide with calendar years. With a distribution of shifts between covered and noncovered employment spread throughout the year, one would expect an effect in the years before and after any spell of at least a year out of covered employment. Hence two more dummy variables are defined to measure this effect. For a year one year after a zero, we define shock one, S_1 and for a year one year before a zero, anticipatory shock AS:

$$S^h_{1t} = \begin{cases} 1 & \text{if earnings of } h \text{ are zero in year } t-1 \\ 0 & \text{otherwise} \end{cases}$$

$$AS^h_t = \begin{cases} 1 & \text{if earnings of } h \text{ are zero in year } t+1 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

One can now add these two additional variables to the basic regression equation (3) or (6). Since the information is available, the importance of zeros in earlier years is also examined. The formulation allows just one shock for the most recent past zero year. Some tests to allow for several recent zeros produced fairly similar results. Defining 5 shock variables for past zeros we have:[20]

For $i = 1, 2, 3, 4, 5$:

$$S^h_{it} = \begin{cases} 1 & \text{if earnings of } h \text{ are zero in year } t-i \\ & \text{and positive in all years from} \\ & t-i+1 \text{ to } t. \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Thus 6 different shock variables are included in the model:

$$W^h_t - \bar{W}^h = \sum_{j=1}^{9} b'_j (A'^h_{jt} - \bar{A}'^h_j) + \sum_{i=61}^{64} b_i (A^h_{it} - \bar{A}^h_i) + \sum_{k=1}^{5} d_k (S^h_{kt} - \bar{S}^h_k) + e(AS^h_t - \overline{AS}^h) + u^h_t \tag{9}$$

Before proceeding to the fitted equations including these additional dummies, let us identify the data problem associated with the anticipatory shock variable and discuss the two methods employed to deal with the problem. For an employee whose earnings from a particular employer exceed the taxable maximum, the data tape contains an estimate of annual earnings. The estimate is constructed by extrapolation to the remainder of the year of the earnings in the quarters before the quarter in which the maximum is reached.[21] A measurement problem naturally arises for an individual who ceases working in covered employment (or changes employers) after reaching the taxable maximum. One signal of individuals who may have ceased working during a year is the absence of any covered earnings in the following year. Thus there are two problems—

[18] For purposes of analysis of the effects of zeros, years of death, disability, or retirement are not considered to be zeros even though their values are excluded from the estimation.

[19] It is estimated that approximately 90 percent of paid employment is covered (Table 27, *Annual Statistical Supplement, 1973, Social Security Bulletin).*

[20] Although estimated earnings are only available starting in 1956, actual earnings up to the taxable maximum are available starting in 1937. Thus there were no problems with use of these dummies for all years.

[21] A single number was used each year for workers reaching the maximum in the first quarter.

estimated earnings are too high whenever a man stops working after reaching the taxable maximum in a year and this situation is far more likely to occur in a year preceding a year with a zero. Thus simply fitting the model as described would give a coefficient for anticipatory shock which is strongly biased toward zero (since the effect to be measured is set to zero by the data construction process for a large fraction of workers)[22] and the combination of mismeasurement of earnings and a biased coefficient may bias the estimates of other coefficients.

One procedure[23] to obtain unbiased estimates of the other coefficients is to eliminate from the data set all observations coming before a year of zero earnings. Results of using this procedure are described in table 5. Of course, no estimate of the coefficient for anticipatory shock can be obtained in this way. This procedure was suggested to us too late to redo the analysis of residuals, which is therefore based on the procedure to be described next. Fortunately, the coefficient estimates do not differ by a great deal between the two procedures.

The alternative procedure is to define anticipatory shock as only being present when a worker is below the taxable maximum; that is, when the measurement error is not present:

$$AS_t'^h = \begin{cases} 1 & \text{if earnings are zero in year } t+1 \text{ and} \\ & \quad \text{below the taxable maximum in year } t \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

This procedure also results in a biased estimate of the coefficient, with the bias being away from zero (i.e., towards a larger estimated decline in earnings from this effect). The problem is that the subset of individuals with zeros in t+1 who are included in the measurement for AS′ is not a random sample. Rather the set includes those with low earnings in year t. Those with larger effects from anticipatory shock are more likely to be included in the sample, i.e., more likely to have low earnings. Thus the coefficient will be biased away from zero. Since the other coefficient estimates are similar under the different formulations of the model to deal with this problem, it was felt to be appropriate to adopt the hypothesis that remaining biases are small.

6. Coefficient Estimates

The details of the coefficient estimates (apart from the individual constants) appear in tables 3–5. For ease of discussion, table 2 contains the coefficient estimates in ratio terms,[24] without the statistical details. The typical lifetime path of wage-indexed earnings is also shown in Figure 1. Before considering the particular coefficients, we can consider statistical significance and goodness of fit. By the conventional t-test, for the larger sample almost all the coefficients are extremely significant.[25] The reported goodness of fit for the explanation of deviations of individual earnings from individual means is small although the standard error of estimate is reasonable. Since the purpose of the model is to simulate lifetime histories, the vastly greater coefficient of closeness of fit that would appear from considering the entire equation (including individual constants) is not really relevant.[26] The equation demonstrates that there is a

[22] A worker who reaches the maximum in the second quarter and then leaves covered employment will be recorded as having four times his first quarter earnings (assuming they exceed his second quarter earnings). Thus there would be no measured decline in earnings as a result of his departure from covered employment.

[23] This procedure was suggested by Franklin Fisher.

[24] Table 2 was obtained by raising e to the power of the coefficients in Table 3 for the column with all variables, S_1–S_5 and AS′ (i.e., taking the natural antilogarithm).

[25] The coefficients measure log earnings relative to those of a 50-year-old. Thus the t statistic tests the hypothesis that individuals of a particular age are distinguishable from 50-year-olds, the coefficient for a 50-year-old having been set to zero.

[26] We are interested in the explanation of variations in a typical individual's history, not in explaining the differences in income level across people by dummy variables.

significant average age-structure to individual earnings which does explain some considerable fraction of the variation in earnings over all lifetimes, while leaving a considerable degree of randomness in earnings which will also be a major component of the simulation to be described below. In addition, the shock dummies also explain a good deal of the variation in deviations from individual means.

Examining the coefficients on the age variables in the different equations, there are several conclusions to be drawn. [27]

[27] Note that the same wage index was used in all the regressions reported in Tables 3-5.

Fig. 1 Ratio of wage indexed earnings to earnings at age 50

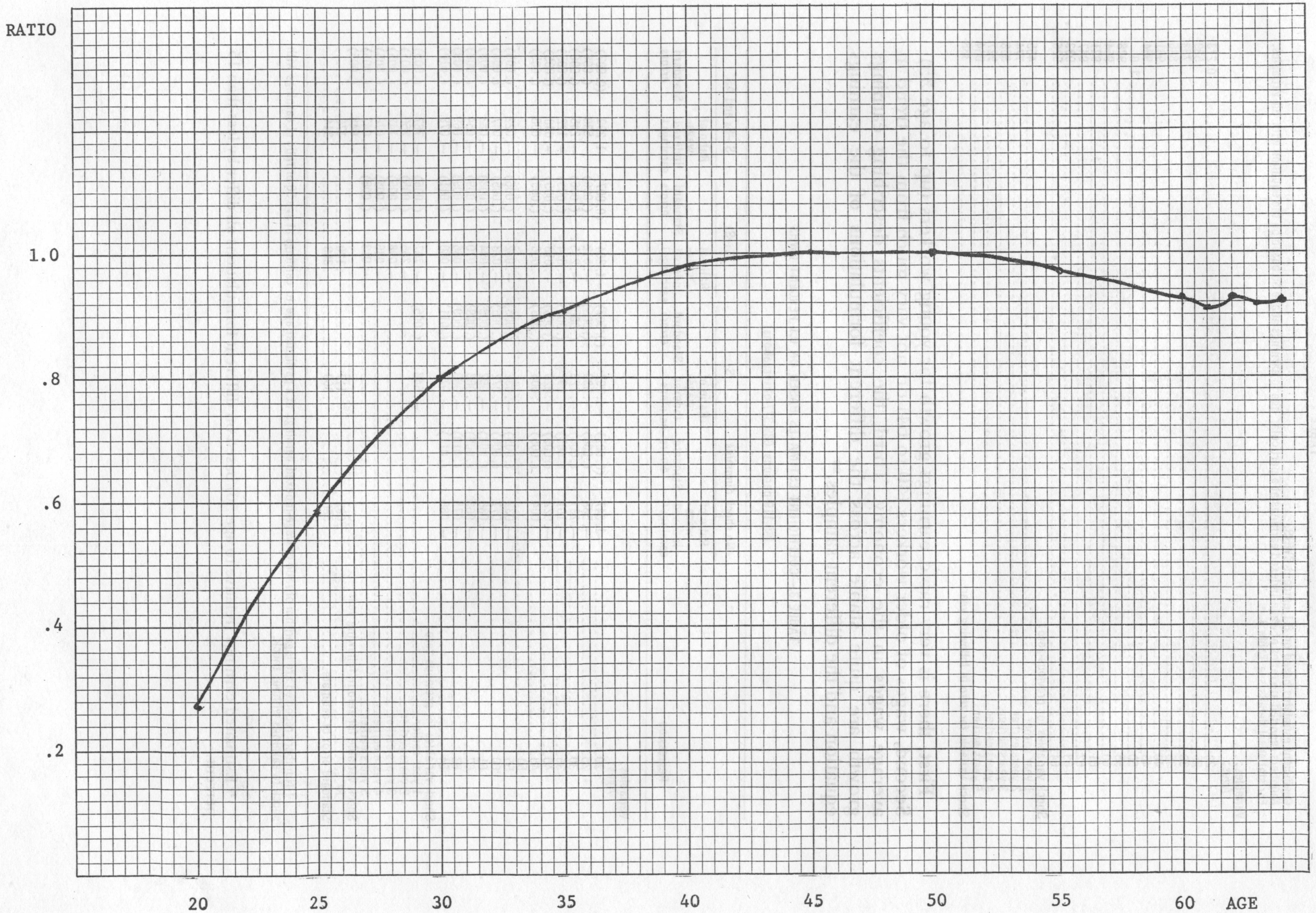


TABLE 2

Coefficient estimates based on 0.1 percent CWHS.
Numbers reported are ratios of wage-indexed earnings at a particular age to earnings at age 50 based on the equation with all variables.
For statistical details see Table 3.

Variables:	
Ages:	
20	0.273
25	.589
30	.783
35	.886
40	.961
45	.994
50	1.0
55	.978
60	.954
61	.921
62	.947
63	.933
64	.926
Shock variables for previous zero:	
1 year earlier	.455
2 years earlier	.774
3 years earlier	.868
4 years earlier	.927
5 years earlier	.954
Shock variable for zero in following year	.388

First, there is very rapid earnings growth for young workers (up to age 35). Second, wages of older workers (40 to 64) do not vary much from the trend in average wages in the economy. Third, the coefficients describing earnings growth are quite stable across the different formulations of the earnings equation and the different samples.[28]

TABLE 3.—COEFFICIENT ESTIMATES BASED ON 0.1 PERCENT CWHS

[65,119 persons, 689,377 observations]

Regression	No shock dummies: Ratio of wage-indexed earnings [1]	No shock dummies: Standard errors	S_1: Ratio of wage-indexed earnings [1]	S_1: Standard errors	S_1–S_5: Ratio of wage-indexed earnings [1]	S_1–S_5: Standard errors	S_1–S_5 and AS': Ratio of wage-indexed earnings [1]	S_1–S_5 and AS': Standard errors
Variables:								
Ages:								
20	−1.347	(0.007)	−1.297	(0.007)	−1.264	(0.007)	−1.299	(0.007)
25	−.588	(.006)	−.539	(.006)	−.512	(.006)	−.530	(.006)
30	−.237	(.006)	−.226	(.006)	−.215	(.006)	−.245	(.006)
35	−.101	(.005)	−.097	(.005)	−.094	(.005)	−.122	(.005)
40	−.022	(.005)	−.020	(.005)	−.018	(.005)	−.040	(.005)
45	.004	(.005)	.005	(.005)	.006	(.005)	−.006	(.005)
50								
55	−.032	(.005)	−.032	(.005)	−.032	(.005)	−.022	(.005)
60	−.071	(.006)	−.072	(.006)	−.073	(.006)	−.047	(.006)
61	−.088	(.008)	−.092	(.008)	−.094	(.008)	−.083	(.008)
62	−.063	(.009)	−.068	(.009)	−.070	(.009)	−.054	(.008)
63	−.075	(.010)	−.083	(.009)	−.086	(.009)	−.069	(.009)
64	−.066	(.011)	−.075	(.011)	−.080	(.011)	−.078	(.010)
Shock variables for previous zero:								
1 year earlier			−.648	(.005)	−.707	(.005)	−.767	(.005)
2 years earlier					−.205	(.005)	−.256	(.005)
3 years earlier					−.098	(.005)	−.142	(.005)
4 years earlier					−.036	(.005)	−.076	(.005)
5 years earlier					−.013	(.006)	−.047	(.005)
Shock variable for zero in following year							−.948	(.005)
R^2	.092		.117		.120		.176	
Standard error of estimate	.380		.369		.368		.344	

[1] Numbers reported are logarithm of ratio of wage-indexed earnings at a particular age to earnings at age 50 or ratio of earnings with shock to earnings without shock.

[28] Estimated earnings (relative to age 50) based on different formulations all differ by less than 15 percent.

TABLE 4.—COEFFICIENT ESTIMATES BASED ON SMALLER SAMPLES

[1,576 persons, 16,747 observations]

Regression	No shock dummies: Rate of wage-indexed earnings [1]	No shock dummies: Standard errors	S^1: Rate of wage-indexed earnings [1]	S^1: Standard errors	S^1-S^5: Rate of wage-indexed earnings [1]	S^1-S^5: Standard errors	S^1-S^5 and AS': Rate of wage-indexed earnings [1]	S^1-S^5 and AS': Standard errors	S^1 and AS': Rate of wage-indexed earnings [1]	S^1 and AS': Standard errors
Variables:										
Ages:										
20	−1.425	(0.048)	−1.358	(0.047)	−1.312	(0.047)	−1.324	(0.046)	−1.380	(0.046)
25	−.645	(.043)	−.603	(.042)	−.564	(.043)	−.567	(.042)	−.615	(.041)
30	−.337	(.040)	−.311	(.039)	−.289	(.039)	−.308	(.038)	−.334	(.038)
35	−.157	(.037)	−.138	(.037)	−.128	(.036)	−.141	(.036)	−.153	(.036)
40	−.057	(.033)	−.048	(.033)	−.038	(.032)	−.049	(.032)	−.061	(.032)
45	−.023	(.034)	−.010	(.033)	−.005	(.033)	−.007	(.032)	−.012	(.032)
50										
55	−.044	(.036)	−.038	(.035)	−.035	(.035)	−.025	(.034)	−.028	(.035)
60	−.021	(.042)	−.025	(.041)	−.024	(.041)	.006	(.040)	.004	(.040)
61	−.070	(.057)	−.086	(.056)	−.089	(.056)	−.080	(.054)	−.077	(.054)
62	−.071	(.063)	−.066	(.062)	−.076	(.062)	−.059	(.060)	−.048	(.061)
63	−.082	(.069)	−.086	(.068)	−.093	(.068)	−.084	(.066)	−.075	(.066)
64	−.055	(.074)	−.062	(.073)	−.070	(.072)	−.058	(.070)	−.048	(.071)
Shock variables for previous zero:										
1 year earlier			−.685	(.030)	−.778	(.032)	−.831	(.031)	−.716	(.029)
2 years earlier					−.288	(.033)	−.333	(.032)		
3 years earlier					−.162	(.033)	−.202	(.033)		
4 years earlier					−.102	(.035)	−.141	(.034)		
5 years earlier					−.032	(.036)	−.064	(.035)		
Shock variable for zero in following year:							−.827	(.029)	−.807	(.029)
R^2	.090		.120		.125		.169		.162	
Standard error of estimate	.414		.401		.398		.378		.381	

[1] Numbers reported are logarithm of ratio of wage-indexed earnings at a particular age to earnings at age 50 or ratio of earnings with shock to earnings without shock.

TABLE 5.—COEFFICIENT ESTIMATES BASED ON SMALLER SAMPLE EXCLUDING YEARS BEFORE ZEROS

[1,528 persons, 16,010 observations]

Regression	No shock dummies: Ratio of wage-indexed earnings [1]	No shock dummies: Standard errors	S_1: Ratio of wage-indexed earnings [1]	S_1: Standard errors	S_1–S_5: Ratio of wage-indexed earnings [1]	S_1–S_5: Standard errors
Variables:						
Ages:						
20	−1.454	(0.048)	−1.376	(0.044)	−1.307	(0.044)
25	−.656	(.040)	−.605	(.040)	−.548	(.041)
30	−.347	(.038)	−.317	(.037)	−.283	(.037)
35	−.159	(.035)	−.140	(.034)	−.124	(.034)
40	−.042	(.031)	−.028	(.030)	−.015	(.030)
45	−.001	(.031)	.009	(.031)	.016	(.031)
50						
55	−.027	(.034)	−.025	(.033)	−.024	(.033)
60	.012	(.040)	−.001	(.039)	−.002	(.039)
61	−.062	(.052)	−.082	(.051)	−.088	(.051)
62	−.050	(.058)	−.054	(.057)	−.068	(.057)
63	−.079	(.064)	−.091	(.063)	−.103	(.062)
64	−.039	(.068)	−.051	(.067)	−.063	(.066)
Shock variables for previous zero:						
1 year earlier			−.728	(.031)	−.845	(.033)
2 years earlier					−.329	(.032)
3 years earlier					−.169	(.032)
4 years earlier					−.152	(.033)
5 years earlier					−.085	(.034)
R^2	.109		.143		.150	
Standard error of estimate	.345		.332		.329	

[1] Numbers reported are logarithm of ratio of wage-indexed earnings at a particular age to earnings at age 50 or ratio of earnings with shock to earnings without shock.

The only curious numbers in the tables are the dips in age 61 earnings relative to neighboring years. Recalling that age 62 is the minimum age for early retirement, one would expect workers who are doing poorly relative to their own life histories to be more likely to collect social security benefits [29] and so be excluded from the sample for analysis. Thus it seems reasonable to conclude

[29] From information on early retirees, it is true that those with low lifetime records do, on average, retire earlier. This is a somewhat different proposition from the speculation in the text.

that there would be a noticeable decline in typical earnings trajectories if early retirement were not an available option.[30]

Considering the coefficient estimates on the shock dummies for zeros in the recent past we have a somewhat different picture. The estimates do vary somewhat across formulations, although not enormously.[31] There is the curious puzzle of the systematic and large (relative to the conventional standard error estimate) differences between the equations fitted to the smaller sample and those fitted to the entire 0.1 percent CWHS.[32] We shall argue that the numbers are in the range of plausible values, so it does seem appropriate to base the simulation on the estimates from the larger sample, adjusting arbitrarily for the bias in the estimate of the coefficient on anticipatory shock. If there were no effect of past zeros other than the carry-over of noncovered employment into the year after a zero, and if switches to covered employment were uniformly distributed over a year, the coefficient of shock one would be one-half. There are three complications to add to this argument. First, there is probably a strong seasonal pattern to job switching. Given the suspicion that moves are concentrated in the late spring and early fall (with more in the former), the seasonal pattern may not affect the argument greatly. Second, there is a complication even if all job switches were uniformly distributed over the year. If the distribution of lengths of time out of covered employment were the same for all dates of switching, the fraction of switches coming after a period out of covered employment which includes an entire calendar year would decrease with the time of the year.

Thus, on average the coefficient on shock one should represent an earnings decrease of less than 50 percent. Third, switching probably lowers earnings[33] (at least in part since some switchers are coming from unemployment or nonparticipation in the labor force) implying a coefficient larger than one-half in absolute value. From these considerations, the estimates of shock one seem to be in a plausible range. The other coefficients for the effects of past zeros show a steady decline in the effects of a previous absence from covered employment, as one would expect.

The estimate of the effect of a zero in the year following a particular year seems too large. Comparing the coefficient with that of shock one, the above argument based on a uniform distribution in the timing of job switches, works in the same way. The seasonal pattern probably makes the effect of anticipatory shock larger. The relationship between switching and earnings is probably weaker. Thus it seems reasonable to expect that the decline in earnings for anticipatory shock is roughly the same as that from shock one. As was discussed above there are reasons to think that this estimate is biased away from zero. In future estimation it would be interesting to develop alternative procedures to obtain an unbiased estimate of this coefficient. The movement in and out of covered employment is sufficiently slow that the exact parameter values on the effects of a zero are not critical components in cost estimation.

To test the robustness of the procedure, the same model was fitted with two modifications. One is the use of a different wage index-the average male earnings of 25- to 64-year olds in the smaller sample. (The index is shown in table 1.) Comparing first and last years, the new wage series shows 5.14 percent growth per year over the period, while the series used above shows 4.48 percent growth per year. This difference of 0.66 percent per year is important for comparing the two regression results. The second modification is to eliminate all observations on 20-24 year olds. Since many of these workers may have been in school and may have had covered earnings from part-time jobs, their

30 For further discussion of this point see section 10.

31 Estimated earnings based on different formulations vary up to 20 percent.

32 The fact that the distribution of the residuals is very far from normal might play a role in explaining the magnitude in differences, but not the persistent sign of the difference in coefficients. Possibly relevant is the fact that the smaller sample was not randomly selected from the CWHS.

33 This expectation is consistent with significant coefficients for earlier shocks.

inclusion might be affecting the other age coefficient estimates by affecting the estimates of a^h, the individual constants. The results are reported in table 6 including a regression using the wage series for those 20 and over for the sake of comparison.

TABLE 6.—COEFFICIENTS WITH ALTERNATIVE WAGE INDEX

Regression	Index of mean wage, 25 to 64 excluding 20-to-24-year-olds		Index of mean wage, 25 to 64			Index of mean wage, 20 and over	
	Ratio of wage-indexed earnings [1]	Standard errors	Ratio of wage-indexed earnings [1]	Standard errors	Adjusted coefficients	Ratio of wage-indexed earnings [1]	Standard errors
Variables:							
Ages:							
20			−1.110	(.046)	−1.308	−1.324	(0.046)
25	−0.415	(0.042)	−.388	(.042)	−.553	−.567	(.042)
30	−.155	(.036)	−.167	(.038)	−.299	−.308	(.038)
35	−.028	(.034)	−.035	(.036)	−.134	−.141	(.036)
40	.025	(.030)	.022	(.032)	.044	−.049	(.032)
45	.031	(.030)	.028	(.032)	−.005	−.007	(.032)
50							
55	−.060	(.033)	−.061	(.034)	−.028	−.025	(.034)
60	−.067	(.038)	−.066	(.040)	−.000	−.006	(.040)
61	−.161	(.051)	−.159	(.054)	−.086	−.080	(.054)
62	−.146	(.057)	−.143	(.060)	−.064	−.059	(.060)
63	−.177	(.063)	−.174	(.066)	−.088	−.084	(.066)
64	−.159	(.067)	−.155	(.071)	−.063	−.058	(.070)
Shock variables for previous zero:							
1 year earlier	−.891	(.034)	−.832	(.031)		−.831	(.031)
2 years earlier	−.419	(.035)	−.332	(.032)		−.333	(.032)
3 years earlier	−.273	(.036)	−.202	(.033)		−.202	(.033)
4 years earlier	−.167	(.036)	−.140	(.034)		−.141	(.034)
5 years earlier	−.095	(.037)	−.062	(.035)		−.064	(.035)
Shock variable for zero in following year	−.795	(.030)	−.825	(.029)		−.827	(.029)
R^2	.113		.157			.169	
Standard error or estimate	.338		.377			.378	
Persons	1,369		1,576			1,576	
Observations	14,235		16,747			16,747	

[1] Numbers reported are logarithm of ratio of wage-indexed earnings at a particular age to earnings at age 50 or ratio of earnings with shock to earnings without shock.

The second column contains the regression results using the alternative wage index and ages 20–64. The third column contains the same coefficients adjusted for the difference in wage indexes. The column was constructed by adding to each age coefficient 0.0066 (Age-50). For comparison purposes the fourth column repeats the coefficients reported in table 4 above.

Comparing the latter two columns, one has little difference in the estimates of wage growth resulting from use of these two wage indexes. Column one contains the coefficient estimates when workers aged 20–24 are omitted from the sample. Comparing columns one and two we see that the age coefficients from the two regressions are very similar. Thus inclusion of 20–24 year olds is not seriously affecting the estimates of the age structure. However, the coefficients on the shock dummies do change somewhat, suggesting, as one might expect, that zeros have somewhat different meanings for the very young than for older workers. Past zeros are more important for prime workers than for young workers.

7. Individual Constants

Given the parameter estimates described above, estimates of individual constant terms, a^h, are obtained from the basic equation (9) using the fact that the estimated error is zero when all variables are at their individual means. There are several uses of these constants which are of interest. First, one wants the distribution of the constant terms as an integral part of the cost estimation. The cohorts born between 1926 and 1931 were pooled to develop an estimate of the distribution of individual constants. Using the coefficients from the equation with all shock dummies fitted to the complete 0.1 percent CWHS, the a^h were estimated for the 188 members of these cohorts in the smaller sample.[34] The calculated distribution of a^h has the shape shown in Figure 2. It is interesting that the distribution is distinctly different from normal.

[34] No further adjustment was made for the different numbers of observations used to estimate the different a^h.

Figure 2 Density of individual constants
(mean = -.113
variance = .59)

Second, one is interested in the stability of the distribution over successive cohorts. Estimates of a^h for all individuals in the smaller sample were calculated using the equation with all shock dummies fitted to the smaller sample. The means of a^h by cohort were calculated and are shown in table 7.[35]

TABLE 7.—MEAN INDIVIDUAL CONSTANT (a^h) BY COHORT

[a^h is estimated from coefficients from regression in Table 4 with all variables]

Date of birth	Mean	Date of birth	Mean	Date of birth	Mean
1893	−0.600	1912	−.221	1932	−.021
1894	−.555	1913	−.285	1933	−.140
1895	−.112	1914	−.353	1934	.057
1896	−.003	1915	−.146	1935	−.135
1897	−.636	1916	−.202	1936	−.116
1898	−.739	1917	−.413	1937	−.137
1899	−.395	1918	−.213	1938	.145
1900	−.465	1919	−.501	1939	.153
1901	−.208	1920	−.280	1940	.158
1902	−.488	1921	−.197	1941	.062
1903	−.988	1922	−.108	1942	.176
1904	−.236	1923	−.154	1943	.199
1905	−.424	1924	−.206	1944	.251
1906	−.601	1925	−.037	1945	.011
1907	−.529	1926	.038	1946	.044
1908	−.374	1927	−.219	1947	.104
1909	−.036	1928	−.059	1948	.082
1910	−.639	1929	−.202	1949	.200
1911	−.428	1930	−.168	1950	.040
		1931	.139		

There is a distinct positive trend in these means indicating that later cohorts have, on average, higher earnings paths relative to the rest of the economy than do earlier cohorts.[36] While one might identify many differences between cohorts and differences in the underlying economy[37] which would justify such a trend, any such discussion would be purely speculative in the absence of further analysis of earnings determination. The trend does not appear so large as to vitiate the use of a single model and single distribution of individual constants for cost estimation, although it might be an improvement to examine[38] the determinants of a^h (using a body of data with more individual information) and to extrapolate the pattern into the future.

Third, the estimates of individual constants can be used to test whether the age profiles of earnings are the same for different earnings levels. For this purpose the equation and sample omitting years before zeros was employed. The a^h in each cohort were divided into thirds representing high, medium, and low levels. Then the earnings records of all individuals who had a^h in the top one-third of their cohorts were combined to form a single sample. The basic equation was fitted to this sample. The same procedure was followed for low and middle thirds. The estimates for these three equations are shown in Table 8. The age structures are graphed in Figure 3.

[35] Since observations per person and residuals per person both decrease with earnings level, a weighted mean would have produced biased estimates of the mean a^h in a cohort.

[36] No test has been made of the statistical significance of this trend.

[37] For example, the shift in the age structure of the male labor force will affect the economy-wide mean earnings series.

[38] In his Ph.D. dissertation, Roger Gordon has examined some of the factors affecting a^h, using the Michigan Panel Study data.

Fig. 3 Ratio of wage indexed earnings to earnings at age 50 for different income levels

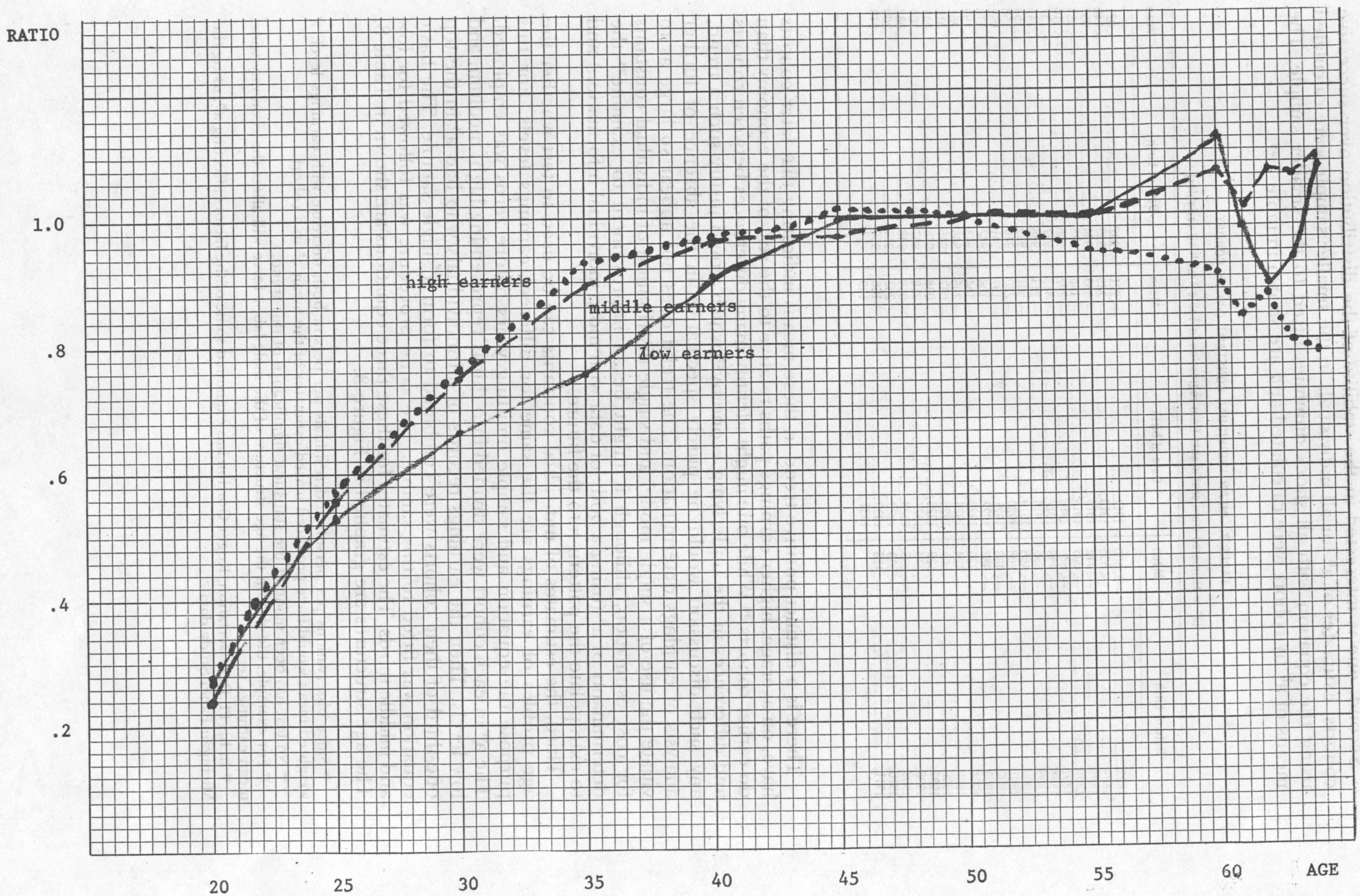


TABLE 8.—COEFFICIENT ESTIMATES FOR DIFFERENT INCOME GROUPS

Regression	Low 3d		Middle 3d		High 3d		Entire population	
	Ratio of wage-indexed earnings [1]	Standard errors	Ratio of wage-indexed earnings [1]	Standard errors	Ratio of wage-indexed earnings [1]	Standard errors	Ratio of wage-indexed earnings [1]	Standard errors
Variables:								
Ages:								
20	−1.310	(0.137)	−1.427	(0.059)	−1.294	(0.045)	−1.324	(0.046)
25	−.634	(.126)	−.584	(.054)	−.557	(.040)	−.567	(.042)
30	−.408	(.117)	−.283	(.049)	−.273	(.037)	−.308	(.038)
35	−.280	(.109)	−.118	(.046)	−.068	(.034)	−.141	(.036)
40	−.098	(.098)	−.043	(.041)	−.030	(.030)	−.049	(.032)
45	−.005	(.099)	−.028	(.041)	.010	(.031)	−.007	(.032)
50								
55	.001	(.106)	.001	(.044)	−.059	(.034)	−.025	(.034)
60	.118	(.126)	.067	(.052)	−.098	(.038)	.006	(.040)
61	−.019	(.072)	.007	(.069)	−.176	(.052)	−.080	(.054)
62	−.119	(.215)	.071	(.077)	−.133	(.054)	−.059	(.060)
63	−.068	(.262)	.059	(.081)	−.226	(.060)	−.084	(.066)
64	.074	(.311)	.093	(.084)	−.244	(.063)	−.058	(.070)
Shock variables for previous zero:								
1 year earlier	−.941	(.065)	−.707	(.051)	−.709	(.041)	−.831	(.031)
2 years earlier	−.447	(.069)	−.166	(.048)	−.293	(.040)	−.333	(.032)
3 years earlier	−.355	(.073)	−.024	(.047)	−.120	(.040)	−.202	(.033)
4 years earlier	−.282	(.078)	.017	(.048)	−.089	(.040)	−.141	(.034)
5 years earlier	−.168	(.084)	.030	(.049)	−.007	(.041)	−.064	(.035)
Shock variable for zero in following year	−.776	(.056)	−.864	(.049)	−1.039	(.047)	−.827	(.029)
R^2	.114		.225		.325		.169	
Standard error of estimate	.899		.237		.130		.378	
Persons	505		566		505		1,576	
Observations	4,581		6,319		5,847		16,747	

[1] Numbers reported are logarithm of ratio of wage-indexed earnings at a particular age to earnings at age 50 or ratio of earnings with shock to earnings without shock.

There are a number of aspects of these equations which are interesting to note.[39] Even before consideration of the coefficients, we can examine the numbers of observations per person appearing in each third of the income distribution.[40] In the lower third, there were 9.1 observations per person; in the middle third, 11.2; and in the upper third, 11.6. Thus zeros are more likely to occur for low income persons. Examining the standard error of estimates in the three equations, we see that the higher the income level the lower the error in estimation. There are two obvious sources for this result—that high income people have less individual noise about their trends and that differences in trends are more important for low earners than high earners (e.g., that the lower third contains a greater fraction of irregular workers who don't have typical earnings paths). Both hypotheses seem plausible.

To compare the age structure of earnings by thirds of the income distribution, we can examine Figure 3. The paths, of course, are roughly similar. However, there are two surprises in the diagram, relative to our expectations. First, it is the high earners who have relative earnings declines as they approach retirement age. While this can be thought of as a natural consequence of a higher income elasticity of the demand for leisure at these ages than when younger (which does not seem implausible), it runs counter to the expectation that low earners would experience far more difficulty in maintaining earnings. However, at later ages the difference might be due to a greater tendency to retire (and thus leave the sample) for lower earners experiencing earnings declines than for higher earners with similar experiences. The second surprise occurs in consideration of earnings when workers are in their thirties. High and middle earners approach their lifetime maxima more rapidly than do low earners. Put differently, high and middle earners experience more of their wage growth at younger ages than do low earners. This runs counter to an image of low earners getting close to their peaks at far younger ages than high earners.

Considering the coefficients on the shock variables, past zeros are considerably more important for the lowest earners than for the other two groups. In the absence of data on the reasons for zeros, one can only speculate that this might

[39] By the Chow test, the equations differ significantly from each other at the 1 percent level.

[40] The lower quality of the estimate of a^h when there are fewer observations might tend to move a somewhat higher fraction of those with fewer observations into both upper and lower thirds.

reflect a greater frequency of job shifts out of covered employment for high earners and a greater frequency of moves out of the labor force for low earners, with the implied differences in work experience and health. The coefficient for anticipatory shock gets larger the higher the income level. Given the bias away from zero in that coefficient arising from the taxable maximum, one would have greater bias the higher the income level of the group.

Since the results reported in Appendix A confirm the view that growth paths are similar by income level, except at the bottom, further work in this area might explore a basis for eliminating very low earners from the sample.

8. Residuals

There are a number of questions about the residuals which are of interest. Of course one wants to know their size and pattern, especially since the simulation depends in an important way on the shape of the entire distribution and not simply its variance. Further, one would expect a significant age structure to the residuals. It is interesting to examine autocorrelation in the residuals. Examining residuals separately by person, it is interesting to examine the relationship between the size of residuals and the level of individual constant (i.e., earnings path).

Using the equation with all shock dummies fitted to the 0.1 percent CWHS sample, the residuals were calculated for each year and each person in the small sample and adjusted for degrees of freedom for that person.[41] The residuals were then separated by the age of the person in each year, with all residuals for ages 20–28 pooled to calculate a density function. The same procedure was followed for ages 29–37, 38–46, 47–55, 56–64. The densities were used for the simulation. They are shown in Figure 4. Surprisingly, the estimated distribution of the residuals gets tighter the older the individuals involved. While this is to be expected in moving from the youngest workers, it is surprising to find the distribution continuing to get tighter as one moves to the largest ages considered. Perhaps the latter result is partially a consequence of the elimination of individuals from the sample when they begin receiving retirement benefits since the analysis by income level showed sharply greater variances for low earners than for high earners and retirement at age 62 is disproportionately concentrated among the workers with lowest earnings.

---

[41] The adjustment made was to multiply each residual by the square root of the ratio of the number of observations for that person to the number minus one.

Fig. 4 Distribution of Random Terms

a. Age: 20-28

Mean of Exp.: 1.243
Prob. ($|x| \leq 1.55$) = .954

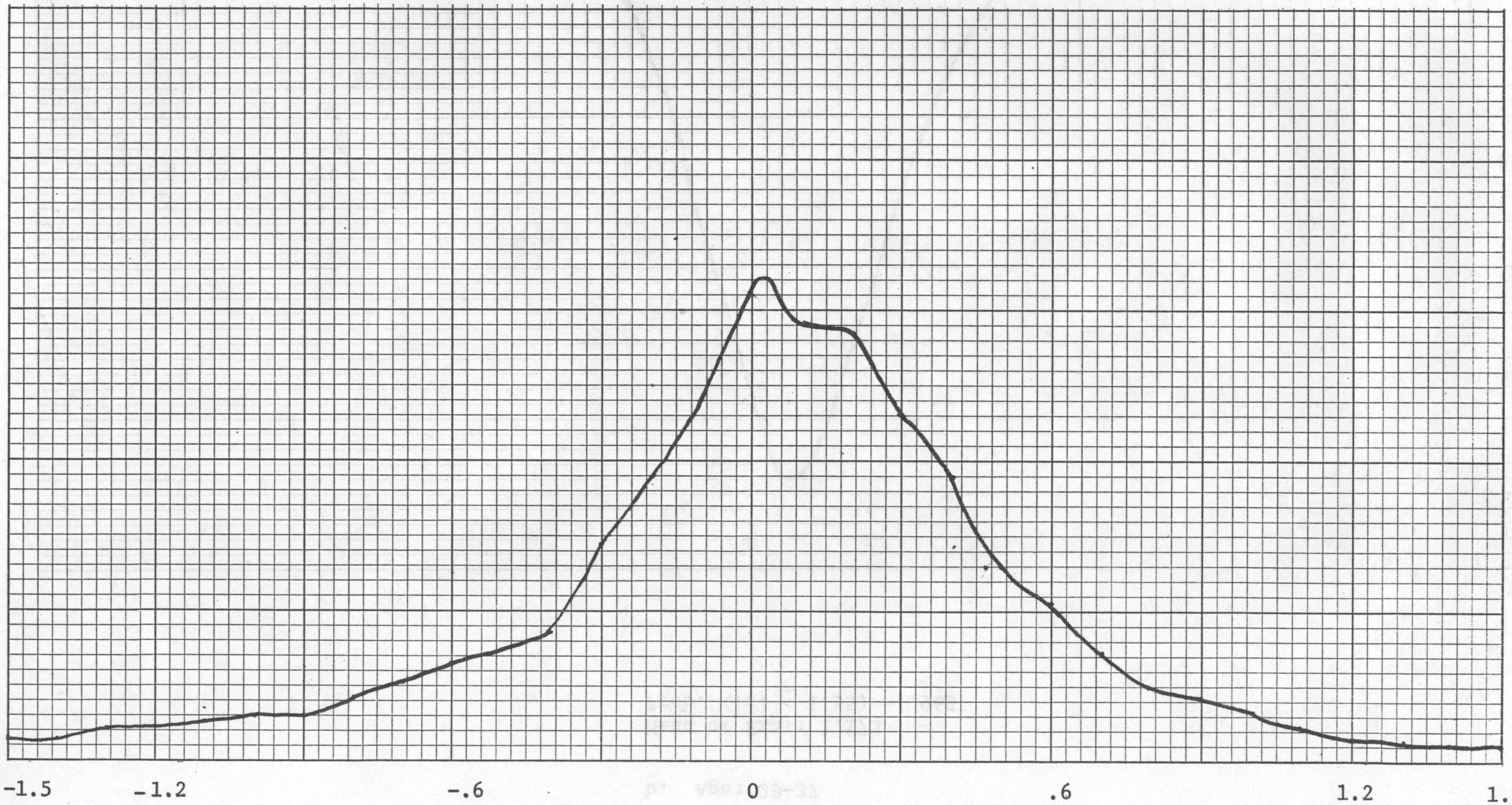


Fig. 4 Distribution of Random Terms (cont'd)

b. Age: 29-37

Mean of Exp.: 1.173
Prob. ($|x| \leq 1.55$) = .963

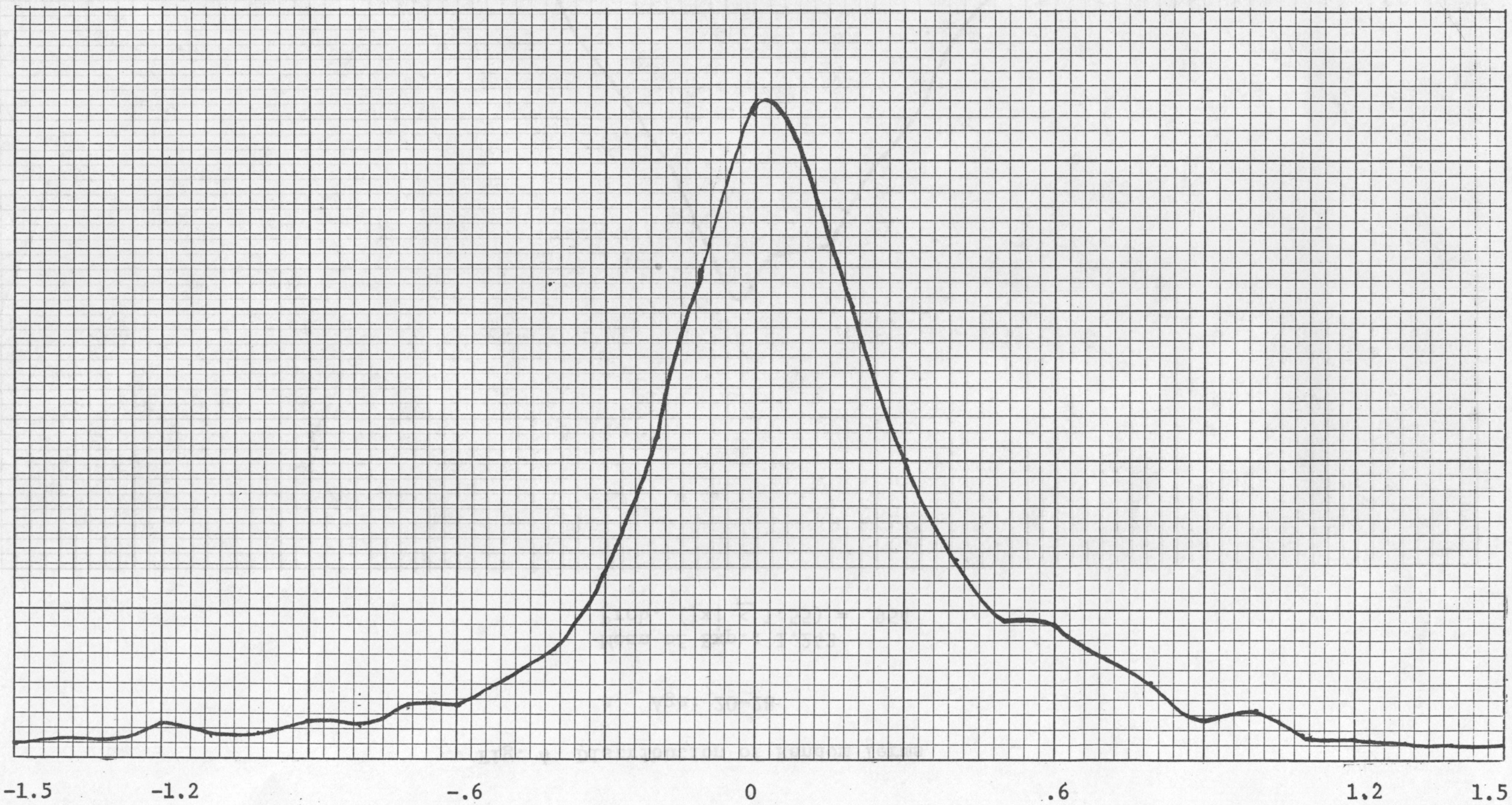


Fig.4 Distribution of Random Terms (cont'd)

c. Age: 38-46

Mean of Exp.: 1.162
Prob. ($|x| \leq 1.55$) = .971

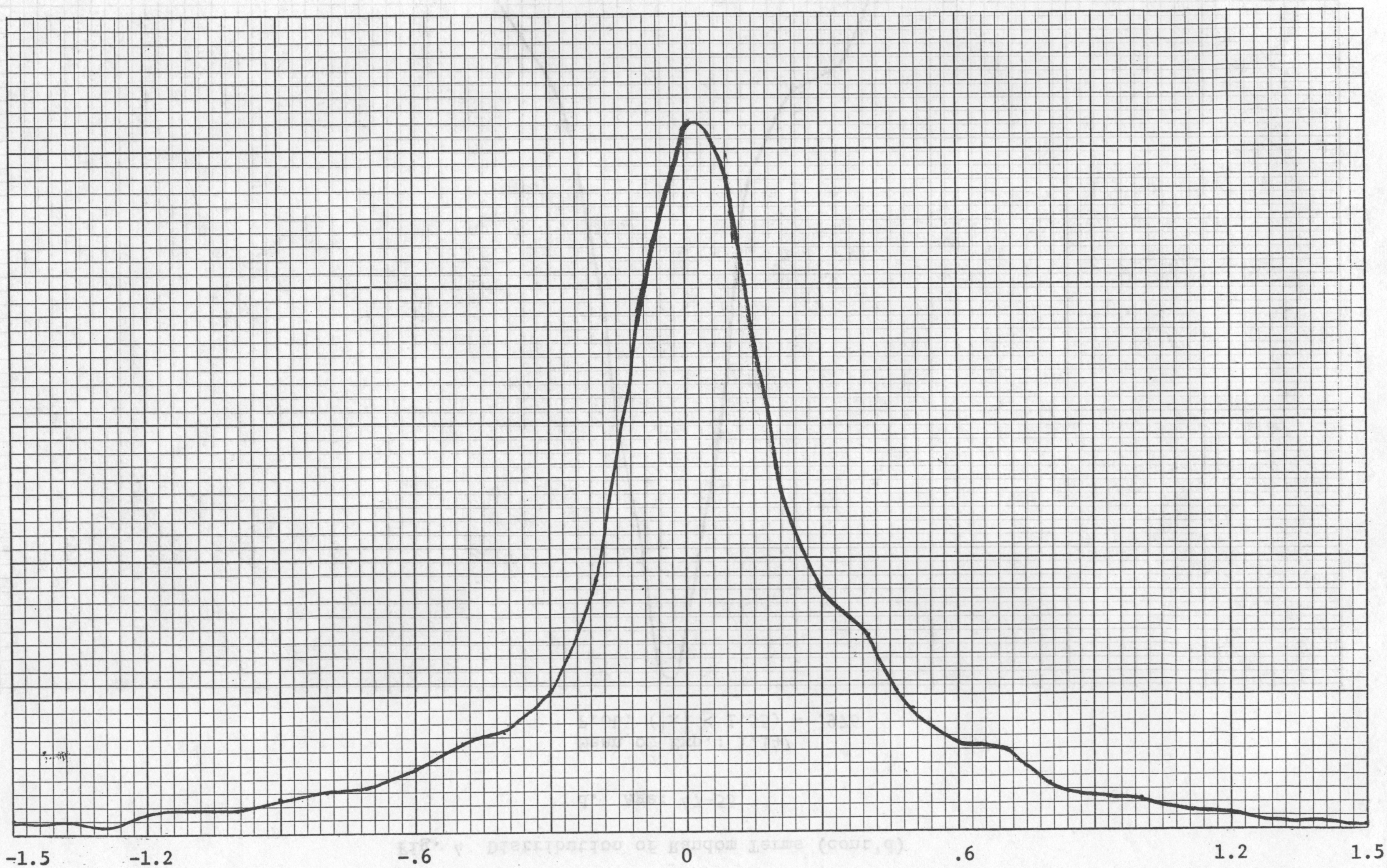


Fig. 4 Distribution of Random Terms (cont'd)

d. Age: 47-55

Mean of Exp.: 1.147
Prob. ($|x| \leq 1.55$) = .976

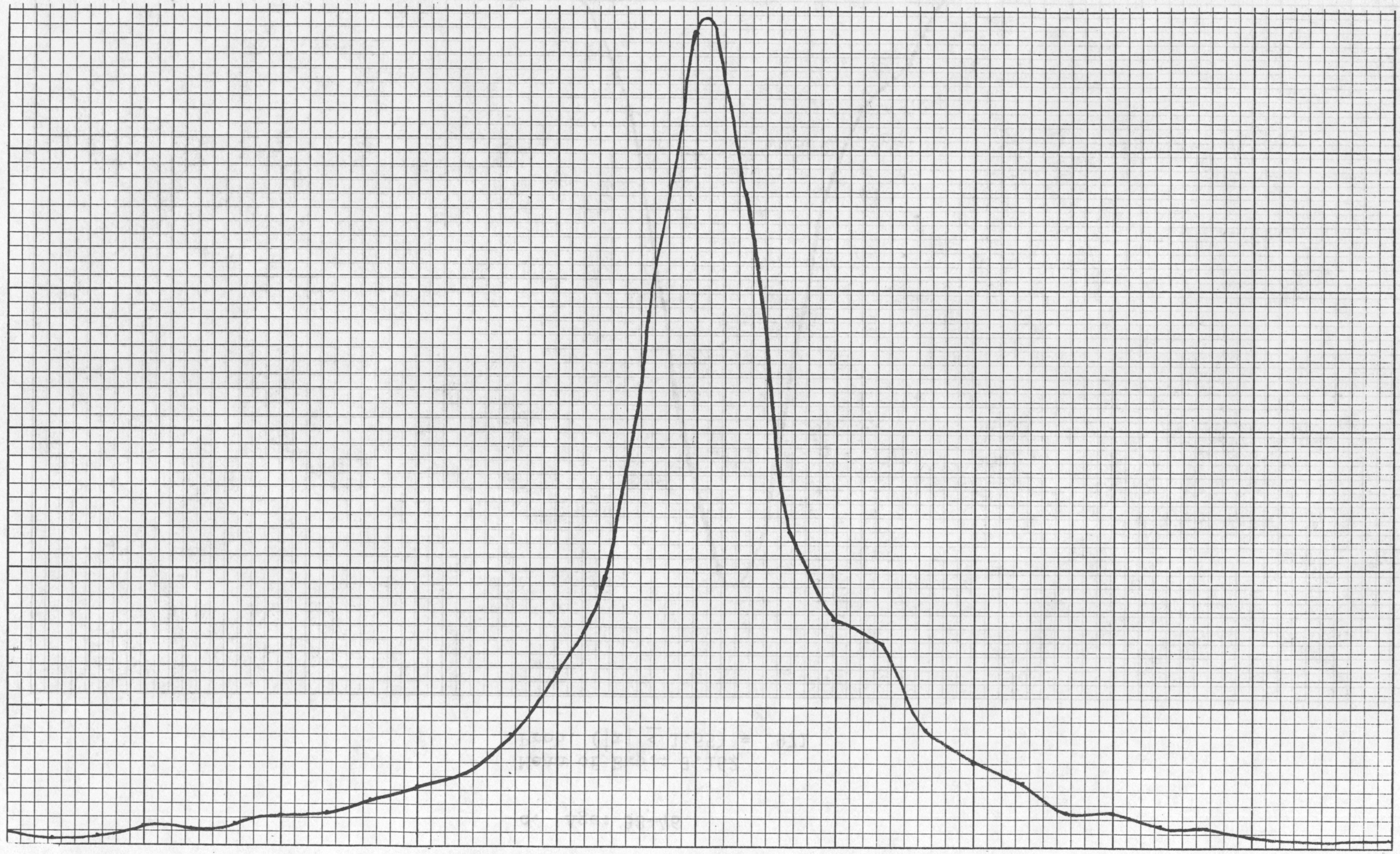
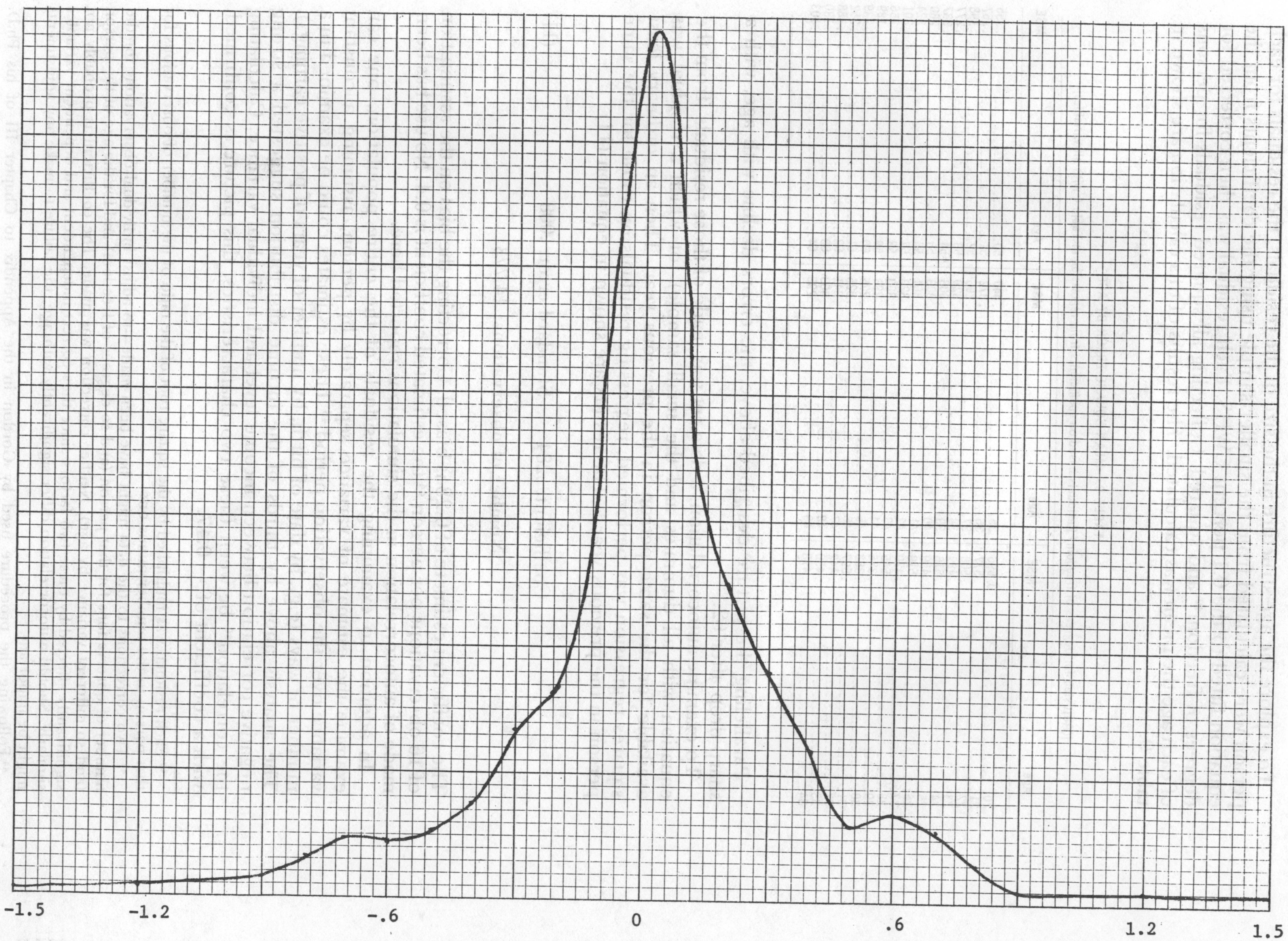


Fig. 4 Distribution of Random Terms (cont'd)

e. Age: 56-64

Mean of Exp.: 1.123
Prob. ($|x| \leq 1.55$) = .983



To explore the age structure in the size of the residuals, the residuals for each person were calculated from the smaller sample using the coefficients from the equation fitted to the 0.1 percent CWHS. Each residual was then corrected for degrees of freedom[42] and squared. Collecting all squared residuals for persons of each age the mean was calculated. The results of this calculation are shown in table 9.

TABLE 9.—MEAN SQUARE ERROR BY AGE

[(Adjusted for degrees of freedom) Residuals based on coefficients from regression in Table 4 with all variables]

Age	MSE	Age	MSE	Age	MSE
20	0.596	35	.440	50	.355
21	.577	36	.372	51	.267
22	.506	37	.296	52	.198
23	.489	38	.329	53	.277
24	.516	39	.308	54	.337
25	.459	40	.303	55	.260
26	.519	41	.361	56	.324
27	.420	42	.323	57	.301
28	.545	43	.363	58	.256
29	.386	44	.293	59	.189
30	.358	45	.275	60	.239
31	.597	46	.325	61	.252
32	.442	47	.341	62	.188
33	.430	48	.403	63	.130
34	.478	49	.298	64	.103

Paralleling the picture described above, the errors decline with age, with a sharp drop after 62.

To examine autocorrelation, a data set was made of those residuals (from the equation with all dummies and the small sample) for which a residual was available for the same person in the previous year. Then an ordinary least squares regression was performed, regressing residuals on those for the same person in the previous year.[43] The results are shown in equation (11)

$$\text{Coefficient: .284} \qquad \text{Standard error: .008} \tag{11}$$

$$\text{Number of observations: 14,773}$$

The coefficient estimate of 0.28 is biased. Correcting the bias on the assumption of 16 observations per person.[44] the estimated coefficient is 0.4. No use has been made of autocorrelation in the simulations reported here.

In addition to examining the residuals of the entire population, one can examine the estimate of variance separately by person, assuming a constant variance over the observation period.[45] These estimates could be examined in a number of ways (e.g., by date of birth, by number of years of positive earnings). The analysis above by thirds of the income distribution suggested a strong negative correlation between income level and individual variance. Calculating the correlation[46] between these two characteristics, this picture is confirmed, with a coefficient of − 0.47.

[42] Each residual was multiplied by the square root of the ratio of the number of observations for that person to the number minus one.

[43] Two corrections to the data might have been made—to adjust individual observations for the number of observations for that person (and so degrees of freedom) and for the age of the person (and so different variance by age). Neither correction was made. The coefficient is probably not significantly affected by these two factors since they correct independent and dependent variables similarly. Should autocorrelation differ significantly with age, the failure to make any adjustments might be important.

[44] Following the procedure used by Gordon in the Appendix to Chapter III of his Ph.D. dissertation.

[45] These estimates are based on the small sample excluding years before zeros.

[46] In doing this calculation no adjustment was made for the variances in the estimation of a^h and individual variance. A weighted correlation might have been different.

9. Individual trends

Just as the basic model described above in equation (3) contains individual constants for the height of the earnings path, it is natural to consider adding an individual trend to the equation. Then (ignoring zeros) the basic equation would take the form:

$$W_t^h = a^h + \Sigma b_i A_{it}^h + c^h(\text{Age}_t^h - 50) \qquad (12)$$

where Age_t^h is the age of person h in year t. Considering entire lifetimes rather than 16-year observation periods it is not completely plausible that an individual will have earnings growth that is consistently more rapid than the typical path in his cohort; that is, the validity of extending a 16-year model to 45 years seems lower for a trend term than for an intercept term. Nevertheless, this model was briefly explored, although not used in the simulation.

The following iterative procedure was followed: Regress W_t^h on age and shock dummies, then take the residuals for each person separately and regress them on time. Provided the distributions of the c^h are the same for every cohort,[47] the omitted variables are independent of the included variables in the first regression. Thus the estimates of the coefficients on age and dummies are unbiased and consistent. Regressing the residuals on time for each person provides an asymptotically unbiased estimate of c^h (asymptotic in the number of persons in the first regression). Since age is nonstochastic, had the sample been constructed to include given numbers of each age, the estimates of c^h would be unbiased too.

Following this procedure, of course, gives the same age structure. One can then examine the distribution of c^h by cohort to test the stability of the model and the validity of the procedure. One can also examine the correlation between a^h and c^h to examine whether high income people have high earnings growth. Starting with the latter question the correlation is 0.02, showing a very weak relationship as was suggested by the separate regressions for different income levels. Table 10 contains the mean value of c^h by cohort.[48]

TABLE 10.—MEAN INDIVIDUAL RATE OF GROWTH BY COHORT (c^h)

Date of birth	Mean	Date of birth	Mean	Date of birth	Mean
1894	0.099	1913	.021	1932	−.003
1895	−.020	1914	−.028	1933	−.006
1896	.024	1915	−.004	1934	−.017
1897	−.009	1916	.053	1935	−.002
1898	−.085	1917	.009	1936	−.015
1899	−.087	1918	.008	1937	.005
1900	−.033	1919	−.019	1938	.025
1901	−.009	1920	.052	1939	.023
1902	−.025	1921	.013	1940	.013
1903	−.053	1922	.003	1941	.048
1904	−.019	1923	.041	1942	−.034
1905	−.009	1924	−.013	1943	.036
1906	.010	1925	−.001	1944	.012
1907	−.032	1926	−.015	1945	−.011
1908	−.037	1927	−.029	1946	.066
1909	−.006	1928	−.003	1947	.004
1910	.024	1929	.017	1948	.013
1911	−.118	1930	−.020	1949	−.068
1912	−.002	1931	.022		

There is a slight trend apparent in these coefficients.[49] This trend implies that under the assumptions underlying equation (12) the procedure will produce biased estimates of the age structure of earnings.[50] By an F test, the entire set of coefficients was found to be statistically significant at the 1 percent level.

[47] A similar condition must hold for shock dummies.

[48] Again, an unweighted mean was used.

[49] No statistical test has been performed on the significance of this trend.

[50] It is interesting to not that with this further adjustment, a repeat of the autocorrelation regression on these residuals yields a much lower coefficient of 0.06.

10. Unemployment Rates

In considering simple way of extending the model, one that comes immediately to mind is the use of some measure of the business cycle as an independent variable. For simulation purposes use of such a model would require simulating future business cycles. Thus, for simulation purposes the business cycle was left as part of the residual noise, leaving an assumption that future business cycles will have comparable severity to those experienced in the observation period. Nevertheless a little analysis was done to include the business cycle to evaluate the changes in parameters induced in this way.

If the dependent variable were real wages or wages relative to some trend, one would expect an increase in the unemployment rate to lower wages. However the effect of higher unemployment on wage-indexed wages is more complicated. One would expect that part of the effect would be an increase in the variance of residuals. Conventional discussions also suggest that unemployment falls more heavily on the younger and older workers, suggesting that their relative wages would decline with high unemployment, while those of medium age workers would therefore increase. To capture this effect we have defined two unemployment-age variables. Both variables are the national unemployment rate or zero depending on the age of the particular worker:

$$U_{1t}^{h} = \begin{cases} U_t & \text{if worker } h \text{ is less than 35 years old in year } t \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

$$U_{3t}^{h} = \begin{cases} U_t & \text{if worker } h \text{ is greater than 54 years old in } t \\ 0 & \text{otherwise} \end{cases}$$

where U_t is the national unemployment rate for males 20 and over.

In table 11 are reported the results of this regression on the small sample.

TABLE 11.—EFFECTS OF UNEMPLOYMENT (16,747 OBSERVATIONS, 1,576 PERSONS)

	Wage-indexed earnings [1]	Standard errors
Variables:		
Ages:		
20	−1.178	(0.057)
25	−.437	(.051)
30	−.181	(.048)
35	−.102	(.037)
40	−.055	(.032)
45	−.000	(.033)
50		
55	−.008	(.042)
60	.032	(.054)
61	−.059	(.062)
62	−.037	(.068)
63	−.063	(.073)
64	−.037	(.077)
Shock variables for previous 0:		
1 yr earlier	−.834	(.031)
2 yr earlier	−.333	(.032)
3 yr earlier	−.203	(.032)
4 yr earlier	−.139	(.034)
5 yr earlier	−.059	(.035)
Shock variable for zero in following year	−.823	(.029)
Unemployment—Age interaction:		
20 to 34	−.025	(.006)
55 to 64	−.006	(.009)
R^2	.170	
Standard error of estimate	.378	

[1] Numbers reported are logarithm of wage-indexed earnings at a particular age to earnings at age 50 or ratio of earnings with shock to earnings without shock.

The effect of an increase of one point in the unemployment rate is a 2.5 percent decline in earnings of a young worker relative to his earnings if he were age 50. The equation also shows a small decline in earnings for older workers, however the coefficient is not statistically different from zero.

11. Mean Earnings by **Age**

The basic model we have explored uses longitudinal data to examine the wage paths of individuals relative to movements in an economy-wide average over the same period. Given an age structure which is assumed to hold over long periods, given a stable distribution of individual constants across cohorts, and given the distribution of residuals,[51] the model implies a pattern to the cross-section structure of mean earnings at any time. We have riot done the calculations to relate the model to the implied cross-section structure. Nevertheless, for the purposes of this project, it is useful to examine the matrix of mean earnings by age, sex, and year to explore the stability of the cross-section pattern of indexed earnings. Table 12 contains this matrix calculated from the 0.1 percent CWHS, where for each sex the earnings in each year have been deflated by average earnings in that year of those included in the tabulation.[52]

51 The distribution of the residuals is relevant since the model has a logarithmic form-the expected value of log wages (which equals the coefficients) will be less than the log of the expected value of wages. This will affect the cross-section pattern since the variance in residuals is not constant across ages.

52 The table reports mean estimated earnings for those between 20 and 65, excluding in a year anyone with zero earnings in that year, anyone who died or received disability benefits in that year, or anyone entitled to old age benefits in that year or earlier. The tabulations for males are based on at least 500 observations, except for ages in the 60's. The female tabulations have about one-half the numbers in the male tabulations. The estimated standard deviation in the estimate of the mean income is generally a few hundred dollars, varying across the matrix.

TABLE 12. MEAN EARNINGS BY AGE, INDEXED BY MEAN EARNINGS BY SEX

	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970	1971
Male:																
21	0.4879	0.4467	0.4484	0.4590	0.4486	0.4274	0.4171	0.4182	0.4335	0.4194	0.3917	0.3703	0.3752	0.3719	0.3761	0.3776
22	.5396	.5357	.4937	.5259	.5222	.4967	.4963	.4909	.4759	.4994	.4918	.4569	.4506	.4587	.4294	.4421
23	.5688	.5521	.5455	.5805	.5873	.5763	.5620	.5816	.5569	.5650	.5819	.5723	.5663	.5598	.5484	.5157
24	.6500	.6290	.5936	.6315	.6113	.6283	.6314	.6232	.6684	.6442	.6611	.6511	.6518	.6507	.6306	.6207
25	.7214	.7670	.6718	.6928	.7061	.6773	.6689	.7034	.6975	.7286	.7094	.7278	.7258	.7167	.7031	.6838
26	.7564	.7843	.7427	.7484	.7737	.7961	.7535	.7402	.7809	.7804	.8176	.7651	.7902	.7847	.7040	.7666
27	.8440	.8321	.8186	.8097	.7939	.8093	.8292	.8140	.8045	.8360	.8195	.8704	.8280	.8535	.8303	.8224
28	.8860	.8867	.8679	.8642	.8560	.8542	.8480	.8621	.8738	.8309	.8942	.8611	.9003	.8787	.9032	.8968
29	.9620	.8425	.9381	.9142	.9145	.9022	.8836	.8806	.8996	.9138	.8982	.9407	.9215	.9553	.9042	.9650
30	.9648	.8856	.9637	.9746	.9526	.9330	.9203	.9757	.9264	.9455	.9619	.9456	.9911	.9579	.9962	.9480
31	1.0012	.9942	1.0319	.9958	.9944	.9957	.9895	.9938	.9507	.9403	.9700	.9981	.9725	1.0230	1.0118	1.0118
32	1.0072	1.0339	1.0534	1.0235	1.0093	1.0121	1.0139	1.0145	1.0017	.9711	.9820	.8988	1.0288	1.0841	1.0757	1.0290
33	1.0331	1.0508	1.0805	1.0535	1.0413	1.0285	1.0270	1.0468	1.0675	1.0107	.9854	1.0238	1.0243	1.0587	1.0370	1.1119
34	1.0524	1.0647	1.0568	1.0884	1.0520	1.0808	1.0511	1.0681	1.0737	1.1018	1.0605	1.0532	1.0731	1.0674	1.0850	1.0755
35	1.0809	1.0681	1.0944	1.0867	1.1420	1.1039	1.1155	1.0861	1.0923	1.0757	1.1160	1.1096	1.0725	1.0940	1.0892	1.1357
36	1.1001	1.1247	1.0731	1.1028	1.0827	1.1199	1.1335	1.1615	1.1116	1.8913	1.1051	1.1339	1.1215	1.1470	1.1323	1.1123
37	1.1331	1.1094	1.1223	1.0978	1.1033	1.0911	1.1387	1.1500	1.1588	1.1045	1.1145	1.1256	1.1739	1.1350	1.1761	1.1583
38	1.1020	1.1778	1.1267	1.1566	1.1511	1.1294	1.0691	1,1281	1.1686	1.1805	1.1291	1.1325	1.1658	1.1332	1.1735	1.1785
39	1.1157	1.1218	1.1680	1.1169	1.1948	1.1393	1.1217	1.1190	1.2090	1.1649	1.2055	1.1429	1.1558	1.1956	1.1985	1.1866
40	1.0828	1.1479	1.1617	1.1786	1.1255	1.1519	1.1449	1.1249	1.1226	1.2233	1.1664	1.2173	1.1506	1.1669	1.1977	1.2295
41	1.1219	1.1169	1.1335	1.1692	1.1890	1.1521	1.1949	1.1531	1.1908	1.1578	1.2287	1.1868	1.2338	1.1623	1.1698	1.2224
42	1.1632	1.1724	1.1252	1.1433	1.1527	1.1895	1.1535	1.1769	1.1439	1.1715	1.2239	1.2398	1.2395	1.2351	1.2107	1.1923
43	1.1803	1.1948	1.1633	1.1256	1.1626	1.1597	1.1874	1.1687	1.1927	1.1836	1.1648	1.1956	1.2634	1.2353	1.2553	1.2624
44	1.1624	1.1674	1.1833	1.1769	1.1089	1.1543	1.1488	1.2169	1.1729	1.3142	1.1828	1.1545	1.2084	1.2682	1.2340	1.2656
45	1.1224	1.1396	1,1931	1.1906	1.1727	1.1083	1.1589	1.1137	1.2315	1.1919	1.2133	1.1819	1.1676	1.2101	1.2787	1.2477
46	1.1236	1.1299	1.1267	1.1835	1.2207	1.1954	1.1157	1.1637	1.1763	1.2225	1.2280	1.2276	1.2068	1.1792	1.2187	1.2660
47	1.0999	1.1665	1.0944	1.1238	1.1884	1.1747	1.1732	1.1166	1.1814	1.1981	1.2274	1.2081	1.2354	1.2095	1.2815	1.2283
48	1.1201	1.1346	1.1480	1.1049	1.1037	1.1741	1.1924	1.1344	1.1174	1.1181	1.1942	1.2201	1.1743	1.2350	1.2191	1.1917
49	1.1270	1.1520	1.1294	1.1295	1.1109	1.0846	1.1595	1.2342	1.1745	1.0984	1.1433	1.1800	1.2224	1.2007	1.2384	1.1938
50	1.1110	1.1490	1.1640	1.0976	1.1031	1.1103	1.0867	1.1539	1.1308	1.1727	1.1792	1.1346	1.1895	1.2307	1.1825	1.2386
51	1.1827	1.0973	1.0967	1.1131	1.1077	1.1820	1.1097	1.0776	1.1722	1.1844	1.1618	1.1058	1.1398	1.1774	1.2369	1.1790
52	1.1912	1.1299	1.1039	1.1184	1.0986	1.1099	1.1483	1.0986	1.0800	1.1387	1.1715	1.1635	1.1091	1.1350	1.1603	1.2547
53	1.0971	1.1629	1.1409	1.0932	1.1303	1.1263	1.2085	1.1388	1.0940	1.0750	1.1271	1.1879	1.1688	1.1245	1.1502	1.1700
54	1.1254	1.8728	1.1525	1.1505	1.0656	1.1170	1.1121	1.0712	1.0895	1.0926	1.0477	1.1356	1.1568	1.1581	1.0651	1.1420
55	1.1062	1.0919	1.1135	1.1530	1.1351	1.0488	1.1155	1.0680	1.0513	1.1097	1.0513	1.0613	1.1092	1.1859	1.1534	1.0655
56	.9935	1.0876	1.1279	1.0965	1.1220	1.1128	1.0226	1.0939	1.0780	1.0585	1.0564	1.0811	1.0881	1.0938	1.1821	1.1303
57	1.1159	.9872	1.0709	1.8514	1.1091	1.1144	1.1316	1.0231	1.0667	1.0741	1.8267	1.1003	1.0807	1.0464	1.0754	1.1439
58	1.0292	1.0883	1.0092	1.0887	1.0478	1.0792	1.1865	1.8912	1.0442	1.0267	1.0830	1.0395	1.0844	1.0510	1.8725	1.9714
59	1.0971	1.8535	1.0525	.9611	1.0286	1.0818	1.0565	1.1055	1.8505	1.0445	1.0245	1.0445	1.0011	1.0899	1.0340	1.0743
60	.8548	1.1270	1.0004	1.0478	.9752	1.0152	1.0170	1.8175	1.1198	1.0492	.9946	1.0422	1.0504	1.0164	1.0498	.9968
61	.8848	.9416	1.0815	.9815	1.0163	.9587	.9669	.5897	1.0332	1.0892	.9968	.9954	1.0408	1.0246	.9663	1.0637
62	.9465	.9784	.9477	1.0983	.9728	1.0784	1.8112	1.1826	1.0660	1.0757	1.1816	1.0875	1.0353	1.1531	1.1505	1.0574
63	.9710	.9382	.9051	.9205	1.0426	1.0879	1.1355	1.1253	1.1377	1.1107	1.0476	1.2540	1.0344	1.0257	1.1784	1.1598
64	.8628	.9715	.9459	.8800	.8981	1.1889	1.1647	1.1286	1.1079	1.2061	1.1661	1.1414	1.2599	1.0791	1.2228	1.1972

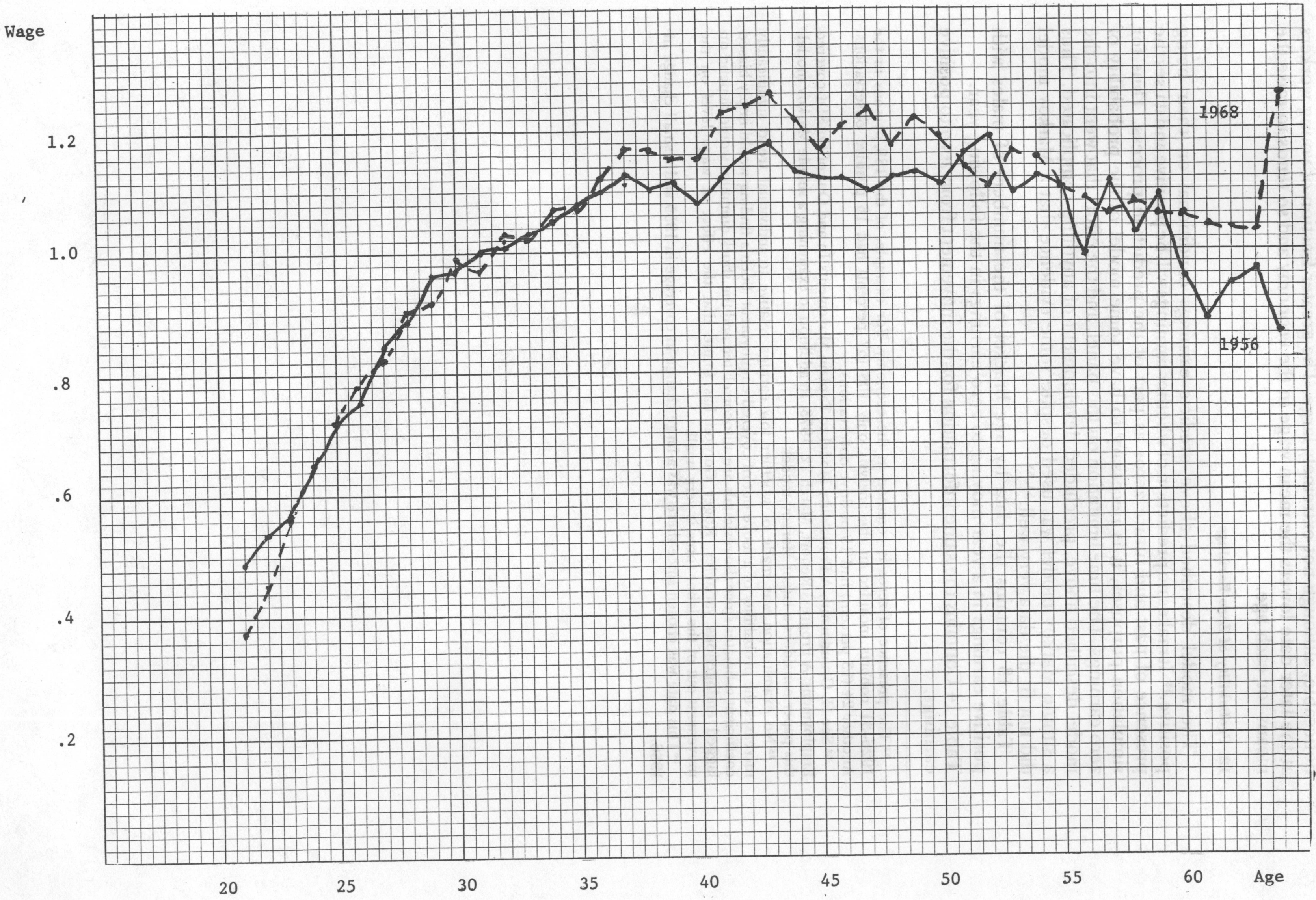
TABLE 12. — MEAN EARNINGS BY AGE, INDEXED BY MEAN EARNINGS BY SEX—Continued

	1956	1957	1958	1959	1960	1961	1962	1963
Female:								
21	.7896	.7957	.7677	.7256	.7470	.7185	.7221	.7390
22	.8510	.8244	.8516	.8422	.7905	.7961	.8044	.8060
23	.8914	.8383	.8627	.8713	.8682	.8316	.8452	.8361
24	.9092	.9006	.8812	.8918	.9010	.8706	.8571	.8743
25	.9171	.8665	.8704	.8881	.9188	.9220	.8909	.8795
26	.9103	.9164	.8883	.8549	.8886	.9156	.9008	.8562
27	.9654	.9006	.8660	.8754	.8407	.8359	.9415	.9036
28	.9323	.9169	.8843	.8940	.8988	.8526	.8625	.8454
29	.8452	.8674	.9176	.8639	.8921	.8518	.8666	.8880
30	.9654	.8581	.9239	.9527	.9668	.8633	.8736	.8984
31	.8757	.8634	.8593	.9195	.9010	.8903	.9197	.9141
32	.9565	.9278	.9586	.8527	.9321	.8757	.8510	.8964
33	.9753	.9421	.9036	.9977	.8651	.9404	.9119	.8723
34	.8888	.9743	.9701	.9181	1.0044	.8800	.9527	.9410
35	1.0168	.8823	.9306	.9459	.9516	.9850	.9098	.9566
36	.9675	1.0183	.9099	.9764	.9796	1.0171	.9886	.9486
37	.9848	.9686	1.0246	.9109	.9590	1.0116	1.0239	1.0177
38	1.0787	.9980	.9470	1.0432	.9072	.9663	1.0054	.9835
39	1.0000	1.0682	1.0101	.9441	1.0226	.9867	.9534	1.0265
40	1.0021	1.0079	1.0829	1.0159	.9951	1.0134	.9769	.9988
41	.9974	1.0410	1.0598	1.0909	1.0732	1.0326	1.0700	.9976
42	1.1007	1.0143	1.0598	1.0000	1.0603	1.0510	1.0156	1.0450
43	1.0509	1.0984	1.0193	1.0759	1.0533	1.0895	1.0531	1.1317
44	1.0656	1.0460	1.1267	1.0109	1.0475	1.1127	1.0953	1.0827
45	1.0855	1.0883	1.0920	1.1000	1.0524	1.0338	1.0852	1.0502
46	1.0756	1.1355	1.0949	1.0932	1.0848	1.0783	1.0650	1.0719
47	1.0992	1.1409	1.1311	1.0850	1.0684	1.0908	1.0247	1.0562
48	1.0908	1.0737	1.1311	1.1423	1.0830	1.0883	1.0902	1.0631
49	1.0672	1.0969	1.1007	1.1187	1.1735	1.0865	1.0795	1.1108
50	1.1427	1.1093	1.0984	1.0955	1.1167	1.1431	1.0848	1.0368
51	1.0740	1.1316	1.1829	1.0900	1.0701	1.1058	1.1866	1.1060
52	1.0771	1.0455	1.1388	1.1887	1.0679	1.1106	1.1363	1.1788
53	1.1317	1.0396	1.0684	1.2451	1.1625	1.0596	1.0877	1.1458
54	1.1301	1.0999	1.0882	1.0873	1.2623	1.1508	1.0478	1.0863
55	1.1957	1.1588	1.1359	1.0713	1.1012	1.2545	1.1655	1.0683
56	1.0435	1.2155	1.1834	1.1237	1.0808	1.0835	1.2026	1.1771
57	1.0950	.9980	1.1860	1.1928	1.1198	1.1195	1.1107	1.2104
58	1.0832	1.1034	1.0251	1.2037	1.1580	1.1063	1.0543	1.1422
59	.9281	1.0529	1.0405	1.0500	1.2175	1.1521	1.0807	1.0241
60	1.0519	.9387	1.0501	1.0373	1.0817	1.1508	1.1350	1.0880
61	.9822	1.1236	.9528	1.0095	1.0346	1.1015	1.1766	1.1631
62	1.2046	1.1192	1.1113	1.0955	1.2277	1.2513	1.2802	1.2610
63	1.1868	1.2043	1.0853	1.1360	1.3305	1.1787	1.2470	1.2735
64	.3722	1.3051	1.2511	1.1623	1.2224	1.2037	1.2715	1.2181

	1964	1965	1966	1967	1968	1969	1970	1971
Female:								
21	.7018	.7168	.7026	.6933	.7166	.6929	.6710	.6376
22	.8280	.7935	.7995	.7694	.7819	.8020	.7805	.7641
23	.8631	.9100	.8905	.9337	.8788	.8814	.8999	.8828
24	.8788	.8865	.9297	.9151	.9515	.9503	.9638	.9363
25	.8891	.8508	.9090	.9590	.9534	.9191	.9404	.9405
26	.9083	.8957	.8314	.9295	.9537	.9846	.9316	.9423
27	.9037	.9247	.9174	.8649	.9000	.9771	.9852	.9521
28	.8835	.9144	.9451	.9285	.8936	.8900	.9356	.9897
29	.9633	.8718	.8331	.9469	.9331	.8257	.9224	.7989
30	.8326	.9735	.8866	.8675	.9750	.8929	.8529	.9290
31	.9293	.9184	.9554	.8304	.9267	.9763	.9462	.8928
32	.8968	.8902	.9129	.9751	.8656	.9243	.9616	.9300
33	.8695	.9060	.8702	.9147	.9868	.8677	.9281	.9332
34	.9823	.8828	.9080	.9164	.9426	.9960	.9106	.9320
35	.8983	.9001	.9272	.9295	.9258	.9297	.9944	.9122
36	1.0340	.9151	.9304	.8983	.9249	.9177	.9520	1.0366
37	.9369	.9934	.9272	.9351	.9144	.9543	.8961	.9536
38	1.0279	.9677	1.0273	.9482	.9469	.9706	.9745	.9739
39	1.0004	1.0287	.9643	1.0272	.9693	.9880	.9713	.9829
40	1.0122	1.0206	1.0738	.9925	1.0414	.9877	.9987	1.0875
41	.9885	1.0140	1.0066	1.0059	1.0037	1.0669	.9987	1.0276
42	1.0237	.9996	1.0453	1.0558	1.0696	1.0043	1.0692	1.0422
43	1.0443	.9893	1.0413	1.0840	1.0755	1.0746	1.0030	1.0668
44	1.0447	1.0470	1.0318	1.0243	1.0647	1.0729	1.0719	.9975
45	1.0593	1.0632	1.0742	1.0361	1.0469	1.0643	1.0963	1.1199
46	1.0891	1.1076	1.0882	1.0879	1.0235	1.0697	1.0580	1.1691
47	1.1135	1.1007	1.0672	1.0974	1.0828	1.0451	1.0899	1.0893
48	1.0830	1.1205	1.1714	1.0964	1.1104	1.1286	1.0695	1.0740
49	1.0478	1.0952	1.1155	1.1217	1.1196	1.1271	1.1136	1.1014
50	1.0501	1.0353	1.1074	1.1174	1.1294	1.1471	1.1023	1.1152
51	1.0586	1.0364	1.0476	1.0938	1.0776	1.1474	1.1294	1.1199
52	1.1372	1.1117	1.0665	1.0669	1.1147	1.1234	1.1820	1.1109
53	1.1774	1.1168	1.1200	1.0777	1.0724	1.1489	1.1347	1.1779
54	1.1433	1.1380	1.1193	1.0931	1.1095	1.0874	1.1318	1.1335
55	1.0692	1.1422	1.1767	1.1250	1.0785	1.1351	1.0993	1.1415
56	1.1208	1.1264	1.1718	1.1571	1.1491	1.1263	1.1345	1.1132
57	1.2229	1.1029	1.0976	1.1702	1.1957	1.1049	1.1396	1.1084
58	1.2167	1.2006	1.1106	1.1282	1.1506	1.1865	1.1248	1.0916
59	1.1506	1.2399	1.2050	1.1876	1.1491	1.1471	1.2169	1.1856
60	1.0483	1.1613	1.1052	1.2530	1.1494	1.1037	1.1475	1.1937
61	1.1323	1.0093	1.1733	1.2270	1.1556	1.1043	1.1020	1.1262
62	1.3326	1.3005	1.1470	1.3914	1.3518	1.2865	1.2568	1.2156
63	1.2856	1.4236	1.2306	1.2155	1.2887	1.3251	1.3720	1.2509
64	1.3151	1.4635	1.4351	1.3339	1.2730	1.2854	1.3374	1.2793

From the perspective of the possible use of wage indexed earnings for benefit calculations, there are three implications of this table worth noting. The large difference in mean earnings between men and women implies that the wage index depends upon the sex mix of the labor force. Over the time period examined the ratio of mean male earnings to mean earnings (both sexes) has varied over a range of 5 percent. Continued growth of female labor force participation rates will see further changes in this ratio. Second, the large differences in mean earnings by age implies that a change in the age structure of the labor force will change the mean wage-indexed wage by age even if the cross-section pattern is unchanged. Third, a change in the mix of experienced and inexperienced workers might alter the age structure of mean earnings. To show these latter two effects, the mean wage by age for male, indexed by mean male wage, is plotted in Figure 5 for 1956 and 1968. The differences are

Fig. 5 Mean Wage-Indexed Wage by Age, Male, 1956 and 1968



noticeable, although not extremely large. The greater fraction of young workers at the later date lowers the mean wage in the economy and so tends to raise the means for each age.

12. Probability of Zero Earnings

The model described above predicts earnings conditional on their being positive.[53] It makes no prediction ofwhether earnings will be positive and it uses the presence of past and future zeros as part of the prediction process. Thus for simulation purposes, it is necessary to have some model of the probability of zero earnings. The time limitation on this panel, and the effort that went into the model described above, precluded development of anything complicated. Thus a simple Markov model was used despite some evidence that a Markov model did not fit the data very well.[54,55]

Table 13 contains the data by age bracket of the numbers of males with positive earnings in a year moving to zero earnings in the following year.

Table 14 Contains the same information for the movement from zero to positive earnings.

[53] The presence of zeros is common in the economy. For example, of the 188 persons in the 1926-31 cohorts included in the analysis only 108 (57 percent) had 16 positive observations. Another 24 (13 percent) had 15 positive observations.

[54] See H. Grundmann, A Probability Model to Explain Movements In and Out of OASDI-covered Employment: A Progress Report, March 22, 1973. In addition to examining a simple Markov model, this paper considers a latent Markov model.

[55] No examination was made of whether low earnigs relative to individual trend significantly increases the probability of a move out of covered employment. Such a finding would imply biased coefficient estimates since the residuals are autocorrelated and the shock dummies would depend on lagged endogenous variables. Roger Gordon has found that the effect of wage level on the movement out of the labor force is very small.

56This date was chosen since there was a large change in coverage in 1955 and a small change in 1956.

TABLE 13.—NUMBERS MOVING FROM POSITIVE EARNINGS IN A YEAR TO ZERO EARNINGS THE FOLLOWING YEAR

	Age																					
	20 to 24		25 to 29		30 to 34		35 to 39		40 to 44		45 to 49		50 to 54		55		56 to 61		62		63		64	
Date	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]	R [1]	N [2]
1951	664	3,456	204	4,292	161	4,028	130	3,434	93	2,681	68	2,066	61	1,436	6	258	31	944	5	87	1	73	3	71
1952	546	3,206	183	4,428	148	4,243	114	3,586	74	2,820	76	2,213	67	1,591	9	216	42	1,111	3	114	4	91	5	74
1953	654	3,223	270	4,486	176	4,256	157	3,803	116	3,028	113	2,372	75	1,703	10	273	61	1,201	10	138	4	120	4	89
1954	362	3,192	160	4,277	126	4,361	96	3,844	71	3,161	78	2,502	43	1,758	10	311	38	1,321	3	118	8	131	9	118
1955	326	3,731	192	4,516	148	4,650	135	4,209	110	3,596	85	2,858	66	2,130	10	306	60	1,621	6	208	5	149	4	148
1956	143	3,993	135	4,501	122	4,778	126	4,442	116	3,820	103	3,033	59	2,344	14	409	59	1,783	4	209	7	218	12	152
1957	206	4,949	176	4,869	157	5,024	141	4,812	111	4,039	99	3,186	78	2,500	12	415	59	1,946	13	256	11	212	10	218
1958	152	4,966	147	4,777	148	5,103	112	4,760	107	4,219	81	3,340	72	2,625	11	432	83	2,070	8	298	9	246	10	205
1959	179	5,113	158	4,859	119	4,993	117	4,949	103	4,302	91	3,520	92	2,805	9	436	88	2,222	8	259	12	295	26	247
1960	185	5,260	192	4,908	134	4,971	147	5,020	125	4,461	111	3,721	87	2,936	11	505	79	2,342	21	300	16	256	22	284
1961	171	5,327	120	4,940	140	4,898	143	5,024	128	4,606	109	3,880	91	3,013	16	556	77	2,514	20	327	19	285	27	244
1962	204	5,752	153	4,995	150	4,893	147	5,002	125	4,778	113	4,015	94	3,161	13	552	82	2,753	25	323	27	312	32	279
1963	175	6,031	129	5,135	121	4,827	104	5,072	113	4,751	114	4,223	79	3,311	19	572	104	2,919	27	393	20	301	34	300
1964	179	6,373	140	5,296	107	4,893	102	4,989	99	4,918	126	4,293	95	3,518	14	608	84	3,065	29	397	42	374	35	295
1965	171	6,584	118	5,456	105	4,964	100	4,994	121	5,029	101	4,441	82	3,704	17	646	89	3,282	30	425	24	381	34	348
1966	209	7,110	163	5,527	135	5,118	126	4,967	136	5,069	100	4,625	127	3,914	10	645	118	3,483	32	440	23	408	40	369
1967	218	7,385	168	5,935	96	5,092	127	4,946	122	5,029	111	4,809	106	3,983	26	725	115	3,598	38	492	47	418	47	398
1968	251	7,664	157	6,252	133	5,233	108	4,896	137	5,146	112	4,765	100	4,183	15	722	108	3,715	48	544	44	465	38	378
1969	284	7,922	241	6,565	152	5,341	135	4,921	132	5,005	150	4,915	119	4,261	27	770	143	3,890	53	533	56	514	52	434
1970	407	8,206	303	6,634	189	5,411	158	4,920	163	4,940	195	4,919	165	4,367	29	808	189	4,025	65	551	54	485	75	469
1971	342	8,115	282	6,937	194	5,399	189	4,944	180	4,802	200	4,852	187	4,431	44	798	247	4,152	59	560	73	493	65	441
Subtotal 1957–71	3,333	96,757	2,647	83,085	2,080	76,160	1,956	74,216	1,893	71,094	1,813	63,504	1,574	52,712	273	9,190	1,665	45,967						
Subtotal ratio (percent)	3.4		3.1		2.7		2.6		2.7		2.9		3.4		3.0		3.6							

1 Number moving to zero earnings in following year.
2 Number with positive earnings in given year by age in that year.

Note: The table is constructed for males who survived to 1972. All those uninsured, entitled to disability benefits in this period or with average earnings per year employed below $500 (during the ages included in the table) were eliminated from the sample of the 0.1 percent CWHS with which the calculations were done.

Given the change in eligibility at age 62, no subtotals were calculated for greater ages. Age 61 should have been excluded also, but wasn't.

TABLE 14.—NUMBERS MOVING FROM ZERO EARNINGS IN A YEAR TO POSITIVE EARNINGS THE FOLLOWING YEAR—Continued

Date	Age: 20 to 24		25 to 29		30 to 34		35 to 39		40 to 44		45 to 49		50 to 54		55		56 to 61		62		63		64	
	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]	R[1]	N[2]
1951	508	1,307	235	816	170	727	128	563	104	381	59	282	35	198	9	37	33	139	4	12	1	9	2	5
1952	625	1,559	227	811	180	742	138	572	108	416	66	311	52	235	9	39	35	153	7	18	2	13	2	9
1953	597	1,543	204	800	139	721	106	599	86	409	63	325	46	264	6	47	22	183	1	16	2	14	3	15
1954	850	1,645	382	882	303	819	265	674	229	471	158	389	136	309	21	44	105	244	14	29	10	25	9	16
1955	565	1,237	283	703	223	653	189	544	123	340	127	300	97	234	24	47	71	188	7	22	4	18	7	23
1956	950	1,084	487	669	345	575	261	505	178	371	85	250	57	222	17	41	52	185	5	25	6	21	3	19
1957	135	245	118	332	103	352	105	377	85	305	76	270	58	246	14	38	57	196	3	34	2	24	3	22
1958	185	331	165	380	138	409	123	395	91	363	82	288	73	288	10	46	57	205	5	29	8	44	6	33
1959	162	299	147	359	126	388	119	404	111	384	95	299	74	289	12	51	59	247	4	31	1	32	3	45
1960	175	323	134	353	120	377	117	404	95	394	76	305	73	309	9	53	72	282	6	42	4	35	3	43
1961	186	335	152	409	123	381	132	422	107	431	125	377	68	312	18	64	72	300	5	46	11	57	8	47
1962	188	327	144	374	92	381	107	433	123	463	95	365	79	333	16	69	83	322	3	45	11	61	10	65
1963	215	377	137	366	121	420	139	490	103	457	101	401	96	353	17	73	69	322	7	65	13	67	6	77
1964	194	352	131	334	127	420	125	438	147	486	122	434	83	333	21	63	93	367	12	65	15	85	6	74
1965	198	339	131	331	143	403	128	408	113	436	112	449	106	357	19	66	75	352	11	62	12	182	11	112
1966	163	317	120	328	93	356	109	367	109	418	92	446	91	372	10	52	79	374	8	56	13	81	11	94
1967	205	362	139	383	110	387	113	381	117	449	81	459	86	412	10	68	66	391	10	74	6	80	13	91
1968	195	357	140	413	98	371	97	400	100	447	94	465	81	453	19	90	90	437	18	87	13	102	13	121
1969	229	416	138	424	88	394	90	434	93	445	88	506	68	475	18	83	68	448	5	93	11	117	14	133
1970	208	438	147	515	113	479	96	487	81	476	81	556	84	533	20	101	77	517	7	98	10	141	13	162
1971	300	559	242	718	161	551	135	561	105	536	127	637	122	644	15	121	111	626	12	113	21	156	18	185
Subtotal 1956–71	3,238	5,377	2,185	6,019	1,756	6,069	1,735	6,401	1,580	6,490	1,447	6,257	1,242	5,709	228	1,038	1,128	5,386						
Subtotal ratio (percent)	60.2		36.3		28.9		27.1		24.3		23.1		21.8		22.0		20.9							

[1] Number moving to positive earnings in the following year.
[2] Number with zero earnings in a year.

Note: The table is constructed for males who survived to 1972. All those uninsured, entitled to disability benefits in this period or with average earnings per year employed below $500 (during the ages included in the table) were eliminated from the sample of the 0.1 percent CWHS with which the calculations were done. Given the change in eligibility at age 62, no subtotals were calculated for greater ages.

For use in the simulation model, the subtotals since 1957[7] were calculated to provide an estimate of the probability of movement.[56] With the change in availability of early retirement for 62-64 year old males, no subtotals were calculated for those ages.

As one would expect, the probability of moving to a zero is u-shaped, troughing in the ages 30-44. The range of probabilities varies very little, being around 3 percent. The probability of movement from a zero to positive earnings is very high for the youngest group and then declines. For those over 30, the probability lies between 20 and 30 percent. Thus the movement probability estimates roughly parallel the relationship between covered and uncovered employment, with roughly 90 percent of employment covered by OASDI.

While no detailed analysis of these probability numbers was employed, a quick examination was made of the explanation of these numbers by means of both the linear and logit probability models, using as explanatory variables time and the unemployment rate. The results were similar for different ages and those for ages 40-49 are reported in table 15.

TABLE 15.-MOVEMENTS IN AND OUT OF COVERED EMPLOYMENT

Age	Constant	Time	Unemployment rate
Probability of movement to zero earnings:			
Linear model:			
40 to 44	0.0012	0.0008	0.0029
Standard error	(.0056)	(.0002)	(.0007)
45 to 49	.0128	.0013	.0051
Standard error	(.0039)	(.0001)	(.0005)
Logistic model:			
40 to 44	-4.5429	.0317	.1071
Standard error	(.1888)	(.0067)	(.0252)
45 to 49	-4.9588	.0421	.1773
Standard error	(.1264)	(.0044)	(.0173)
Probability of movement from zero earrings:			
Linear model:			
40 to 44	.4401	-.0081	-.0178
Standard error	(.0410)	(.0015)	(.0056)
45 to 49	.4081	-.0111	-.0028
Standard error	(.0554)	(.0021)	(.0076)
Logistic model:			
40 to 44	.0097	-.0465	-.1046
Standard error	(.2467)	(.0091)	(.0324)
45 to 49	-.1363	-.0631	-.0287
Standard error	(.354:3)	(.0128)	(.0478)

There is a small positive time trend in the probability of moving to zero over the 15 years in the sample, with an increase of approximately one percentage point in the probability of movement per decade. The unemployment rate enters with a positive coefficient (as one would expect), with a one point rise in the unemployment rate increasing the probability of movement by approximately one-half of one percentage point. For the movement from zero, both the time trend and unemployment rate enter with negative signs. The probability of switching to positive earnings decreases by approximately 1 percentage point per year and by approximately 1.5 percentage points for each one point rise in the unemployment rate.[57]

13. Simulation

The empirical analysis described above has been used as the basis of a stochastic simulation model to examine the effects of economic growth and the lengthening averaging period on the cost estimates. One step in the cost estimation procedure used by the Office of the Actuary employs a set of ratios of the average PIA for newly retired workers in the future to the average PIA for newly retired workers in the base year. The focus of the simulation was to produce this set of ratios under alternative economic assumptions. Of course

[57] The results from the linear and logic models are very similar within the estimation period. One would not want to simply extrapolate these time trends far into the future. Doing so, the two models do give different predictions.

this procedure uses a stochastic description of the past as well as a stochastic description of the future, so would be inappropriate for short-term cost estimation.[58]

The first step in the procedure is to select an age profile of wage indexed earnings for a typical worker, assumed to hold for all cohorts.[59] The profile selected is shown in table 16.

This profile can be combined with any assumed growth in real earnings relative to the wage-index used in the estimation to produce a profile of real earnings.[60] Successive cohorts of 20-year-olds are assumed to have initial real wages which grow at the same rate as the growth in national average real wages. The second step in the procedure is to select the distribution of random elements underlying the generation of wage histories.[61] For this purpose, residuals from the equation based on the 0.1 percent CWHS applied to the small sample of workers were used. Each residual was adjusted for degrees of freedom.[62] All residuals associated with individuals in 9 year age brackets were pooled to form a distribution. These are the distributions shown in Figure 4, above. Given this random structure and the wage profile, 100 wage histories are randomly generated.

Using the transition probabilities shown in table 16, these wage histories are then subjected to probabilities of having zero covered earnings. When a zero occurs, the particular earnings level is set to zero and the neighboring earnings levels are reduced by the factors shown in table 16. This gives the wage histories to be used in calculating benefits.

1. Ratio of wage-indexed earnings to those at age 50:

Age	20	25	30	35	40	45	50	55	60	65
Ratio	0.27	0.59	0.78	0.89	0.96	0.99	1.0	0.98	0.95	0.93

2. Probability of zero earnings at age 20 — 0.1

    Probability of zero-earnings in t given positive earnings in t–1 — .03

    Probability of zero earnings in t given zero earnings in t–1:

    - Age 20 to 24 — 0.40
    - Age 25 to 64 — .75

3. Earnings as a fraction of earnings with no neighboring zeros as a consequence of a zero at t:

t–1	0.50	t+3	.85
t	0	t+4	.90
t+1	.50	t+5	.95
t+2	.80		

It is assumed that at each income level the random pattern of earnings is the same. Thus each of the 100 patterns generated above is assumed to occur at each of 12 earnings levels (corresponding to different individual constants in the regression model). The distribution of individual constants was generated in the regression model and will be used to take a weighted average of PIA, after their computation. However, one further step is needed, the calibration of the distribution of constants to produce the mean estimated covered earnings in the economy. Based on the table of mean earnings by age given above, it was assumed that the mean for 50 year olds was 115 percent of the mean for all males. Then the mean earnings for 50 year olds in 1975 was set equal to $8,939, an estimate of the desired number. This calibration corrects for the difference in wage indexes, the effects of zeros, the distribution of individual constants, and the fact that the error distribution gives a zero expectation for the log wage.

58 In addition to treating business cycles as part of the residual, no adjustment was made for the greater prevalence of years with no covered employment before 1955. A correction for that could be incorporated in the procedure in a straightforward manner.

59 The simulation follows the current procedure of estimating PIA assuming all workers retire at 65.

60 Or it can be combined with any growth of a wage index (relative to the index used in estimation) to produce a profile of wage indexed earnings.

61 No adjustment has been made for autocorrelation of residuals in the results reported here, although incorporation of autocorrelation would be straightforward.

62 The adjustment was to multiply each residual by the square root of the ratio of the number of observations for that person to the number minus one.

Using the growth of initial wages for 20 year olds each of the 1200 histories was shifted to give lifetime patterns for different retirement years. [63] For each year, each of the constructed 1200 wage histories in then converted to taxable earnings histories by applying the appropriate taxable maximum. Then, AIME is calculated for each simulated worker and, using the benefit formula, PIA is calculated for each worker. [64] The PIA's for different income levels were weighted to reproduce the distribution of a(h) approximately described above in Figure 2 [65] and an average PIA calculated. The average PIA's were divided by that of 1975 to get the ratios. Table 17 shows the resulting calculation for 2 percent real growth and the benefit formula recommended above.

TABLE 17

(Ratio of PIA(T) to PIA(1975) all measured in 1975 dollars, price index AIME 2 percent real growth 1975 dollars benefit: 80 percent of lst 2,400, 35 percent of next 4,800, 25 percent of excess/

Year	Amount	Year	Amount	Year	Amount	Year	Amount
1975	1.00000	1976	1.02321	1977	1.04584	1978	1.06802
1979	1.07433	1980	1.08062	1981	1.08689	1982	1.09313
1983	1.09933	1984	1.10549	1985	1.11159	1986	1.11766
1987	1.12252	1988	1.12738	1989	1.13221	1990	1.13703
1991	1.14185	1992	1.14666	1993	1.15145	1994	1.15624
1995	1.17402	1996	1.19160	1997	1.20927	1998	1.22678
1999	1.24412	2000	1.26134	2001	1.27843	2002	1.29539
2003	1.31224	2004	1.32901	2005	1.34565	2006	1.36540
2007	1.38504	2008	1.40458	2009	1.42405	2010	1.44344
2011	1.46274	2012	1.48197	2013	1.50113	2014	1.52023
2015	1.53925	2016	1.55823	2017	1.57714	2018	1.59599
2019	1.61480	2020	1.63356	2021	1.65488	2022	1.67657
2023	1.69864	2024	1.72112	2025	1.74400	2026	1.76733
2027	1.79109	2028	1.81528	2029	1.83991	2030	1.86501
2031	1.89058	2032	1.91664	2033	1.94319	2034	1,97027
2035	1.99787	2036	2.02599	2037	2.05465	2038	2.08389
2039	2.11370	2040	2.14411	2041	2.17511	2042	2.20671
2043	2.23892	2044	2.27177	2045	2.30526	2046	2.33939
2047	2.37417	2048	2.40962	2049	2.44575	2050	2.48255

To examine the sensitivity of the calculation to some changes, the calculation was repeated without probabilities of zero earnings. This produced PIA ratios differing by less than 2 percent and, on average, by considerably less. Decreasing the residuals used in the stochastic simulation by dividing all of them by 1.05 produces no noticeable change.[66] Shifting the distribution of individual constants up or down by 10 percent produces only small changes, on the order of 1 percent. The calculations were repeated for 1.5 percent real growth and for 2 percent real growth until 1971 and 1.5 percent thereafter. These calculations are shown in tables 18 and 19.

[63] The calculation was done separately for years with turning points in the application of the averaging procedure. Linear interpolation of AIME was used for intervening years.

[64] The AIME's generated can be used to examine the importance of the lengthening averaging period. The model shows considerably greater declines in AIME than would be predicted from the same age profile in the absence of random elements in the model.

[65] For simulation purposes, the 32 estimates of a" based on fewer than 10 observations were removed from the distribution, leaving 156 values.

[66] Previous calculations with a normal distribution of errors, rather than the one generated by the regression model, produced sizeable differences in the estimated effect of lengthening the averaging period.

TABLE 18

[Ratio of PIA(T) to PIA (1975) all measured in 1975 dollars. Price index AIME 1.5 percent real growth 1975 dollars benefit: 80 percent of 1st 2,400, 35 percent of next 4,800, 25 percent of excess]

Year	Amount	Year	Amount	Year	Amount	Year	Amount
1975	1.00000	1976	1.02055	1977	1.04067	1978	1.06045
1979	1.06461	1980	1.06874	1981	1.07286	1982	1.07695
1983	1.08102	1984	1.08507	1985	1.08911	1986	1.09311
1987	1.09549	1988	1.09784	1989	1.10019	1990	1.10254
1991	1.10488	1992	1.10721	1993	1.10953	1994	1.11185
1995	1.12578	1996	1.13957	1997	1.15248	1998	1.16530
1999	1.17806	2000	1.19072	2001	1.20330	2002	1.21580
2003	1.22823	2004	1.24060	2005	1.25288	2006	1.26634
2007	1.27973	2008	1.29305	2009	1 30629	2010	1.31945
2011	1.33255	2012	1.34562	2013	1.35864	2014	1.37163
2015	1.38459	2016	1.39751	2017	1 41039	2018	1.42323
2019	1.43604	2020	1.44882	2021	1,46720	2022	1.48590
2023	1.50490	2024	1.52421	2025	1,54386	2026	1.56384
2027	1.58417	2028	1.60485	2029	1.62588	2030	1.64731
2031	1.66914	2032	1.69137	2033	1.71402	2034	1.73708
2035	1.76056	2036	1.78449	2037	1.80886	2038	1,83369
2039	1.85899	2040	1.88479	2041	1.91108	2042	1.93787
2043	1.96518	2044	1.99303	2045	2.02144	2046	2.05040
2047	2.07994	2048	2.11005	2049	2.14074	2050	2.17202

TABLE 19

[Ratio PIA (T) to PIA (1975) all measured in 1975 dollars, price index ATME 2 percent real growth until 1971, 1.5 percent thereafter 1975 dollar benefit: 80 percent 1st 2,400, 35 percent of next 4,800, 25 percent of excess]

Year	Amount	Year	Amount	Year	Amount	Year	Amount
1975	1.00000	1976	1.02231	1977	1.04409	1978	1.06548
1979	1,07051	1980	1.07552	1981	1.08052	1982	1.08549
1983	1.09041	1984	1.09532	1985	1.10017	1986	1.10500
1987	1.10801	1988	1,11101	1989	1.11399	1990	1.11697
1991	1.11994	1992	1.12290	1993	1.12586	1994	1.12881
1995	1.14385	1996	1.15874	1997	1.17264	1998	1.18644
1999	1.20014	2000	1.21376	2001	1.22728	2002	1.24069
2003	1.25404	2004	1.26732	2005	1.28050	2006	1.29438
2007	1.30821	2008	1.32196	2009	1.33563	2010	1.34921
2011	1.36273	2012	1.37621	2013	1.38965	2014	1.40305
2015	1.41642	2016	1.42975	2017	1.44304	2018	1.45628
2019	1.46949	2020	1.48268	2021	1.50150	2022	1.52063
2023	1.54007	2024	1.55984	2025	1.57994	2026	1.60040
2027	1.62119	2028	1.64236	2029	1.66389	2030	1.68581
2031	1.70815	2032	1.73090	2033	1.75408	2034	1.77768
2035	1.80171	2036	1.82619	2037	1.85113	2038	1.87654
2039	1.90244	2040	1.92884	2041	1.95574	2042	1.98316
2043	2.01111	2044	2.03961	2045	2.06868	2046	2.09833
2047	2.12855	2048	2.15936	2049	2.19077	2050	2.22278

As expected, changes in the rate of real earnings growth produces sizeable cost estimate differences, as can be seen by comparing tables 17 and 18 with 1.02^t and 1.015^t respectively. From this comparison, one has the estimate that lowering real growth by 0.5 percent results in costs approximately 15 percent higher in 50 years.

The main purpose of this simulation was to demonstrate the feasibility of a straightforward procedure which could be included in the cost estimation procedure of the Office of the Actuary. Clearly, this line of research needs considerable refinement and sensitivity testing before it can play a major role in cost estimation.

14. Covariance of Earnings

The matrix of variances and covariances of earnings at different ages has been computed to examine the acceptability of the assumptions underlying the basic model.[67] These computations indicate that the underlying assumption of an individual constant which does not change over an entire lifetime is not fully acceptable. Rather, the finding of a slow but continued decline in correlation coefficients with larger differences in ages indicates random movements in earnings which are not adequately captured by a first order autocorrelation

[67]This computation was completed after submission of the report but before printing. Thus this section was added to the end of the paper.

process. To pursue this direction of development in the model one could introduce random changes in individual constants. Thus there could be two random components—one leading to a permanent change in individual constant, the other having only a transitory effect on earnings.[68]

The precise calculations reported in table 20 can be described as follows. Consider the set of men who have positive earnings in the data set in each of the two years when they are i and j years old. For each such man and year calculate the log of earnings indexed by average wages in that year. Then for the set of men with positive earnings when they are i and j years old calculate the correlation coefficient between indexed earnings at age i and indexed earnings at age j. Table 20 reports these correlations for particular ages.[69] To read the table one adds the age difference of a row to the age heading a column. Thus the correlation between earnings at 43 and 42 is .80; between earnings at 53 and 51, .73.

TABLE 20.—COEFFICIENTS OF CORRELATION BETWEEN WAGE-INDEXED EARNINGS AT DIFFERENT AGES

Age	33	43	53	Age	33	43	53
Age Difference				Age Difference			
−1	.75	.80	.79	1	.75	.80	.80
−2	.67	.72	.73	2	.70	.72	.75
−3	.60	.68	.70	3	.64	.68	.70
−4	.54	.66	.67	4	.61	.67	.67
−5	.50	.63	.64	5	.57	.63	.64
−6	.46	.59	.61	6	.56	.62	.62
−7	.38	.58	.61	7	.55	.59	.58
−8	.32	.54	.58	8	.53	.58	.56
−9	.29	.52	.54	9	.51	.58	.54
−10	.25	.50	.55	10	.50	.55	.47
−11	.20	.48	.54	11	.49	.53	.45
−12	.17	.45	.51	12	.44	.52	----------
−13	.15	.40	.51	13	.45	.46	----------
−14	----------	.40	.50	14	.45	.48	----------
−15	----------	.39	.44	15	.43	.41	----------
−16	----------	.31	.50	16	.45	.38	----------

As one would expect, the larger the difference in ages, the lower the correlation. In addition the correlations are considerably lower where the younger age is below 30.[70] Analyzing the model showed a variance in individual constants of approximately .6 and of the random component at prime ages of .3 to .4. The autocorrelation was estimated[71] to be .4. If the model in equations (4) and (6) held exactly, i.e., ignoring the effects of zeros, the correlation coefficients would be .76 with a one year gap as opposed to a range of .75–.80 in table 20. With the model, the correlation coefficients would be .68–.66 with a two year gap; .65–.63 with a three year gap; and would decrease to approximately .63–.60 as the gap increased. Considering table 20, the empirical findings are suggestive of a somewhat higher autocorrelation in the short run, together with a lower autocorrelation over longer periods. This is strongly suggestive of a drift in individual constants.

[68]A model of this type can be suggested by James Mirrlees.

[69]In constructing the table, there were approximately (500) (17–|age difference|) observations for the calculation of a correlation coefficient with any given difference between the two ages.

[70]Since the random components were largest at younger ages, this finding was partially to be expected, but with the magnitudes shown in the table.

[71] No attempt was made to estimate autocorrelation separately by age.